

Estimación de parámetros de la covarianza del pronóstico usando máxima-verosimilitud en sistemas dinámicos geofísicos

Nievas Lio Estefanía

Director: Manuel Pulido

Tesis presentada para obtener el grado de
Magister en Estadística Aplicada



Universidad Nacional de Córdoba
Argentina



Estimación de parámetros de la covarianza del pronóstico usando máxima-verosimilitud en sistemas dinámicos geofísicos by Nievas Lio Estefanía is licensed under a [Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Estimación de parámetros de la covarianza del pronóstico usando máxima-verosimilitud en sistemas dinámicos geofísicos

Nievas Lio Estefanía

Resumen

En el filtro de Kalman por ensambles la subestimación del error del pronóstico, debido al tamaño limitado de los ensambles y al error de modelo, es un desafío que requiere de métodos ad-hoc para su tratamiento. Esta deficiencia es comúnmente tratada a través de inflación multiplicativa o aditiva de la covarianza de la predicción.

En general, existen propuestas de estimación de la inflación multiplicativa a través de métodos adaptativos para cada ciclo. Miyoshi et al. propone una estimación adaptativa para la estimación del factor de inflación. En esta tesis se propone el desarrollo de una técnica que permita la estimación de factores multiplicativos de la covarianza, tanto para el caso de inflación multiplicativa como la aditiva, a partir de observaciones distribuidas en el tiempo.

Las metodologías se basan en la búsqueda de los parámetros óptimos que maximicen la verosimilitud de las observaciones. Se proponen y evalúan diferentes metodologías: la optimización de la verosimilitud aproximada, por medio de evaluación en grilla y el método Nelder-Mead, y extensiones del algoritmo EM (Expectation–Maximization).

Las técnicas se evalúan a través de experimentos gemelos en los cuales se generan observaciones sintéticas con los sistemas dinámicos de Lorenz 63 y 96, simulando entornos con y sin error de modelo.

Se obtuvo como resultado la convergencia de los algoritmos desarrollados. Por un lado, la optimización de la función de log-verosimilitud aproximada por medio de los métodos Grid Search y Nelder-Mead proveen estimadores precisos para los parámetros de la covarianza. Por otro, las extensiones del algoritmo EM desarrolladas logran buenas estimaciones en contextos de regímenes casi lineales; sin embargo, estas últimas subestiman o sobreestiman los parámetros a medida que aumenta la presencia de error sistemático en contextos de modelo imperfecto.

Palabras Claves: Asimilación de datos. Filtro de Kalman por ensambles. Algoritmo EM. Parámetros de la covarianza. Factor de inflación.

Índice general

1. Introducción	4
2. Marco Teórico	9
2.1. Modelo de Markov escondido	9
2.2. Métodos y conceptos de Asimilación de Datos	11
2.2.1. Conceptos	11
2.2.2. Teoría de estimación	12
2.2.3. Asimilación variacional de 3 dimensiones	15
2.2.4. Filtro de Kalman	16
2.2.5. Asimilación variacional de 4 dimensiones	17
2.2.6. Filtro de Kalman por ensambles	18
2.2.7. Suavizado de Kalman por ensambles	21
2.2.8. Parámetroización de la Covarianza del pronóstico	21
2.3. Estimación de máxima verosimilitud aproximada	22
2.3.1. Aproximación de la función de log-verosimilitud	22
2.3.2. Algoritmo Nelder-Mead	23
2.4. Algoritmo EM para estimación de parámetros	25
2.4.1. Problema de estimación	25
2.4.2. Derivación del algoritmo para el caso Gaussiano	26
2.4.3. Ecuaciones y pseudocódigo para EM-EnKS	30
3. Metodología y primeros experimentos	33
3.1. Modelos de pronóstico de sistemas dinámicos	34
3.1.1. Modelos	34
3.1.2. Configuración de parámetros para los experimentos	36
3.2. Marco experimental	37
3.2.1. Experimentos gemelos - OSSE	37
3.2.2. Descripción de los tipos de experimentos realizados	40
3.3. Experimentos de referencia: motivación para el factor de inflación	42
3.3.1. Experimentos con modelo perfecto	43
3.3.2. Experimentos con modelo imperfecto	44

4. Estimación por máximo-verosimilitud aproximada	55
4.1. Análisis en el marco de modelo perfecto	57
4.1.1. Validación de la técnica	57
4.1.2. Influencia del tamaño muestral	60
4.2. Estimación de parámetros en el marco de Modelo Imperfecto . .	62
4.2.1. Estimación del factor de inflación α	62
4.2.2. Estimación del parámetro de la covarianza β	66
4.2.3. Estimación de ambas fuentes de error a la vez: (α, β) óptimo	68
5. Estimación mediante el algoritmo expectation-maximization EM	71
5.1. Estimación del parámetro de la covarianza β	72
5.1.1. EM y estimador de β	72
5.1.2. Implementación y experimentos	72
5.2. Estimación del factor de inflación α	76
5.2.1. EM secuencial y estimador de α	76
5.2.2. Implementación y experimentos	79
6. Conclusiones	87
7. Anexo	89
7.1. Glosario	89
7.2. Resumen de experimentos	91
8. Bibliografía	94

Capítulo 1

Introducción

Revisión bibliográfica

La asimilación de datos es una técnica que permite combinar estadísticamente la información de mediciones de diferentes instrumentos, tales como mediciones satelitales, in-situ, de radares, etc., con la información de un modelo matemático que describa la evolución del sistema que se está analizando (Kalnay, 2006 [1]). Se busca estimar el estado de un sistema físico complejo como la atmosfera, el océano, el caudal de un río, etc., a través de observaciones parciales y ruidosas.

La asimilación variacional de tres dimensiones combina óptimamente una matriz de covarianza estática (climatología) y la covarianza del error de las mediciones. La optimización, determinación del estado óptimo que minimiza la función de costo, se realiza a través de técnicas de optimización variacionales (Bouttier et al., 1999 [2]). Una extensión es la asimilación variacional de 4-dimensiones en la cual se busca optimizar una función de costo definida en un intervalo de tiempo (ventana de asimilación), dentro de la cual se distribuyen las observaciones (Lorenc, 2003 [3]). Esta técnica permite considerar la evolución de la matriz de covarianza pero solo dentro de la ventana de asimilación.

La técnica del filtro de Kalman (Kalman, 1960 [4]) realiza la combinación de las observaciones con un modelo dinámico teniendo en cuenta los errores de ambos y determina la media y la covarianza del análisis, bajo la hipótesis de Gaussianidad de los errores del modelo y las observaciones, y asumiendo que el modelo dinámico es lineal. En 1994, Evensen propuso una alternativa a esta técnica que consiste en realizar representaciones de Monte Carlo de las distribuciones estadísticas de los errores a través de partículas o miembros de un ensamble y de esta manera se determina la evolución de la media y la covarianza del sistema, a esta técnica se la denomina filtro de Kalman por ensambles [5]. Con respecto a las hipótesis, en el filtro de Kalman por ensambles el modelo dinámico es asumido no lineal. Y, en muchos casos, la muestra resultante a través de los miembros del ensamble no es necesariamente Gaussiana, aun cuando para la combinación de la información se utilice esta hipótesis.

En general, se observa que aún en condiciones en los cuales no se cumplen con las hipótesis, el filtro de Kalman da resultados razonables y hay convergencia de la técnica. Sin embargo, es necesario evaluar detalladamente la convergencia del filtro. Además, se le suma a esto los problemas de muestreo y el error de modelo, los cuales pueden deteriorar significativamente la calidad del filtro. [6]

Motivación

Cuando se trabaja con modelos meteorológicos, es muy costoso generar una gran cantidad de miembros de ensambles para desarrollar una buena matriz de covarianza a base de ellos, por lo cual se agrava la subestimación, tanto de las varianzas como de las covarianzas. La dimensión del espacio en el que se trabaja tiene en el orden de entre 10^5 y 10^9 variables, y en contraposición se podría generar a partir de un modelo meteorológico un ensamble que tenga en el orden de 50 miembros. Es decir, para estimar matrices de covarianza, con 50 realizaciones del estado del sistema se desea estimar entre $10^5 \times 10^5$ o $10^9 \times 10^9$ parámetros. A esto se le llama problema de pocas muestras. Por lo cual, la covarianza muestral da una estimación grosera con correlaciones espurias y varianzas subestimadas.

Para contrarrestar un poco el problema, cabe aclarar que, aun cuando se trabaja en espacios muy grandes, la dinámica del sistema esta dominada por una cantidad limitada de modos. Los modos son los autovectores del sistema en los cuales se proyecta la solución. Regionalmente los modos relevantes pueden ser del orden de la centena, por lo cual si bien se cuenta con pocos miembros en los ensambles, no son tantas las dimensiones que interesa representar.

El problema de subestimación de la covarianza del pronóstico lleva a sobrevalorar el pronóstico del modelo por encima de las observaciones, resultando como consecuencia que el análisis resultante se asemeje a los pronósticos del ensamble e ignore las observaciones. A esto se le denomina divergencia del filtro.

Antecedentes

Para abordar el problema de subestimación de la covarianza muestral y la consecuente divergencia del filtro en la teoría se proponen dos caminos: localización, e inflación. Con la localización físicamente se asume que las correlaciones más allá de una distancia característica del sistema, son nulas. Por lo tanto, se asume que la matriz de covarianza del filtro está localizada alrededor de la diagonal. Se suele utilizar alguna función localizada, Gaussiana o exponencial. Con esto se logra que la cantidad de parámetros a estimar se reduzca significativamente, aumentando así la proporción muestra sobre parámetros. [7] [8]

Con inflación se aumenta la dispersión de los miembros del ensamble en cada ciclo para evitar el colapso. Se inflan las perturbaciones del ensamble alrededor de la media, y en consecuencia, se inflan las varianzas y las covarianzas cercanas. Sin embargo, también se inflarán las covarianzas espurias.

En el trabajo de Miyoshi (2011) [9] se propone una solución empírica con inflación. Se estima un factor de inflación de manera adaptativa. Es decir, en

cada tiempo, se realiza una estimación proponiendo inferencia Bayesiana del factor de inflación. Se realiza a partir de una igualdad que involucra las matrices de covarianzas del proceso de asimilación [10] que sirven de diagnóstico del error en la estimación de las matrices de covarianzas.

Por otro lado, otra manera de contrarrestar los problemas de trabajar con ensambles, en entornos con error de modelo, es combinar ambas fuentes de error: combinar una covarianza estática con la matriz de covarianzas de los miembros del ensamble. La covarianza estática permite representar las estadísticas de error independientes del tiempo, mientras que con los miembros del ensamble se representan las estadísticas de error dependientes del flujo. Una matriz estática compensa la dificultad computacional de generar miembros del ensamble, y por ende, de generar estadísticas precisas. Además, permite realizar ponderaciones relativas entre el pronóstico y las observaciones, y permite modelar la influencia de las observaciones fuera de los lugares de observación. Un esquema híbrido opera calculando un conjunto de ciclos de asimilación de datos paralelos, con cada miembro del conjunto recibiendo observaciones perturbadas únicas. Este esquema fue propuesta inicialmente por Hamill y Snyder (2000) [11], donde combinan una covarianza estática con la evolución de la matriz de covarianza del estado del modelo dada por un ensamble basado en el filtro de Kalman. Esta combinación la realizan a través de técnicas variacionales.

El interés por estos métodos híbridos es una tendencia que ha sido liderada por los principales centros de pronóstico, entre los que cabe citar Environment Canada (Buehner, 2005), MetOffice UK (Clayton et al., 2013 [12], Lorenc et al. 2015 [13]) el ECWMF (Bonavita et al. 2016 [14]), y el NCEP (Kleist e Ide, 2015 [15]). Por lo que la estimación de como combinar óptimamente la matriz de covarianza de la predicción con la covarianza estática reviste de un gran interés.

Objetivos

En esta tesis se propone desarrollar métodos para la estimación del factor de inflación, pero en vez de estimar un factor para cada paso de asimilación, estimar uno solo para todo los tiempos, es decir, realizar esta estimación a partir del conjunto de observaciones distribuidas en el tiempo. Se busca considerar un gran número de observaciones que permita restringir fehacientemente el parámetro de la covarianza que optimice la verosimilitud. Llamaremos \mathbf{P}^b de covarianza generada por los miembros del ensamble, y α al factor multiplicativo de ella.

Por otro lado, también se buscará estimar a partir del conjunto de observaciones distribuidas en el tiempo, un factor multiplicativo para la matriz de covarianza de error de modelo. Llamaremos \mathbf{Q} a la matriz fija de error aditivo del modelo, y β al factor multiplicativo de ella. Se buscará también estimar ambos parámetros de la covarianza α y β a la vez, quedando representado el error de modelo por la combinación $\alpha\mathbf{P}^b + \beta\mathbf{Q}$.

Objetivo general: desarrollar métodos para la estimación de los parámetros de las covarianzas, α y β , a partir de un conjunto de observaciones distribuidas en el tiempo utilizando metodos de máximo verosimilitud.

Objetivos específicos:

1. Desarrollar una metodología que estime los parámetros de las covarianzas por medio de la optimización de una aproximación de la log-verosimilitud. Realizar esta optimización por medio de las técnicas Grid Search y Nelder-Mead.
2. Desarrollar una metodología que estime los parámetros de las covarianzas por medio de la optimización de la log-verosimilitud explícita. Ampliar el algoritmo EM para realizar esta optimización.
3. Aplicar y evaluar las técnicas desarrolladas en distintos sistemas dinámicos caóticos.

Metodología

Para deducir métodos para la estimación de estos factores se utilizarán en esta tesis el principio de máxima verosimilitud, utilizando como algoritmos de maximización los métodos Grid Search, Nelder-Mead y el algoritmo de Expectation-Maximization.

El método Grid Search es la evaluación en grilla de la medida a optimizar, y el algoritmo Nelder-Mead es un método de optimización iterativo. Este último se eligió por no necesitar gradiente para operar. Sin embargo, para ambos métodos, la función objetivo, la función de verosimilitud, no es calculable directamente. Existen aproximaciones, una de ellas es la desarrollada por Pulido et al. (2017) [16] y es esta la que se optimizará. La aproximación está basada en la hipótesis de Gaussianidad, considerando que se está trabajando con un filtro Gaussiano, las hipótesis serían congruentes entre ambos.

Por otro lado, el algoritmo de Expectation-Maximization (EM) permite estimar parámetros, tanto físicos como estadísticos, que maximicen la función de verosimilitud, sin necesidad de realizar aproximaciones de ella. Para realizar la maximización se utiliza una función intermedia, denominada límite inferior de la evidencia, (ELBO, evidence lower bound). De esta manera se evita la necesidad de calcular la función de verosimilitud que es intratable. Tandeo et al. (2015) [17] desarrollaron un método que combina el algoritmo EM con el filtro de Kalman por ensambles, el método fue utilizado para la estimación de parámetros físicos con un operador de observación altamente no lineal. El método también ha sido utilizado exitosamente para la estimación de parámetros estadísticos, la covarianza del error de modelo, en un sistema caótico no lineal simple (Dreano et al. 2016 [18]). Uno de los objetivos de la presente tesis es lograr extender y aplicar este método para la estimación de los factores multiplicativos de las matrices de covarianza.

El contexto en el que se realiza el trabajo es de modelos numéricos caóticos, particularmente, con un enfoque a los modelos numéricos atmosféricos. Es por esto que las metodologías se probarán en los modelos Lorenz 63 y Lorenz 96, que son modelos de pocas variables, pero que simulan en forma sencilla algunos aspectos de la dinámica atmosférica.

Esta tesis se encuentra organizada de la siguiente manera. En la segunda unidad se describirá el marco teórico de las herramientas utilizadas, describiendo

do métodos de asimilación de datos en general y el algoritmo EM. En la tercera sección se presenta los modelos numéricos que se utilizarán, y la metodología de los experimentos. Además, se muestra en experimentos como el uso de los factores de inflación solucionan problemas de divergencia, lo cual fundamenta el trabajo de la estimación de los mismos. En la cuarta sección se resuelve el problema de encontrar los factores óptimos por medio de la optimización de una aproximación de la log-verosimilitud de las observaciones. Se optimizan ambos parámetros de la covarianza por separado y luego ambos a la vez, tanto con el método Grid Search, como con el Nelder-Mead. Para cerrar, en la unidad 5 se resuelve el problema por medio de la optimización de la log-verosimilitud explícita. Se presentan, a partir de extensiones del algoritmo EM, deducciones de estimadores para ambos parámetros, pseudocódigos de su implementación, experimentos con ambos modelos, y análisis de los mismos. Por último, se presentan conclusiones en la unidad 6, y un anexo en la unidad 7. En este último se puede encontrar un breve resumen de los términos más utilizados en el proceso de asimilación de datos con EnKF, y un resumen de los experimentos realizados a lo largo de esta tesis, con los parámetros seteados en cada caso.

Capítulo 2

Marco Teórico

2.1. Modelo de Markov escondido

En esta tesis vamos a trabajar en el marco de un modelo de Markov escondido. Una amplia introducción y tratamiento de este marco es dada por Cappe et al. (2005) [19].

Un modelo de Markov escondido (HMM en sus siglas en inglés, *Hidden Markov Model*) es definido por un sistema dinámico no-lineal estocástico \mathcal{M}_Ω que evoluciona en el tiempo a las variables no observables (escondidas) $\mathbf{x}_{k-1} \in \mathbb{R}^n$ de la siguiente manera:

$$\mathbf{x}_k = \mathcal{M}_\Omega(\mathbf{x}_{k-1}) + \eta_k, \quad (2.1)$$

donde k representa el índice de tiempo, $\{\mathbf{x}_i\}_{i=1}^K$ son realizaciones de los estados $\{\mathbf{X}_i\}_{i=1}^K$, y ésta es una cadena de Markov. El sistema dinámico depende de un conjunto determinístico y eventualmente estocástico de parámetros denotados por Ω . Se asume además η_k como error aditivo aleatorio de modelo, gaussiano, con media cero, y covarianza $\mathbf{Q}_k = \mathcal{E}(\eta_k \eta_k^T)$, donde \mathcal{E} denota al operador esperanza. Es decir, $\mathbf{X}_k \sim \mathcal{N}(\mathcal{M}_\Omega(\mathbf{x}_{k-1}), \mathbf{Q}_k)$.

Las observaciones en el tiempo t_k , $\mathbf{y}_k \in \mathbb{R}^m$, están relacionadas con las variables ocultas por medio del operador de observación $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, tal que

$$\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \epsilon_k. \quad (2.2)$$

ϵ_k es un error aditivo aleatorio de observaciones, que se asume gaussiano, con media cero, y matriz de error de covarianza $\mathbf{R}_k = \mathcal{E}(\epsilon_k \epsilon_k^T)$, es decir, $\mathbf{Y}_k \sim \mathcal{N}(\mathcal{H}(\mathbf{x}_k), \mathbf{R}_k)$. La representación de errores $\eta = \eta_1, \eta_2, \dots, \eta_K$ es asumida i.i.d al igual que $\epsilon = \epsilon_1, \epsilon_2, \dots, \epsilon_K$. Además, η y ϵ se suponen conjuntamente independientes, e independientes del estado de pronóstico inicial \mathbf{x}_0 , el cual es asumido Gaussiano con media \mathbf{x}_b y covarianza \mathbf{B} , es decir, $\mathbf{X}_0 \sim \mathcal{N}(\mathbf{x}_b, \mathbf{B})$.

El operador de observación es el que mapea el estado escondido al espacio de las observaciones. Como ejemplo cabe citar que la temperatura de la atmósfera se mide a través de refractividad en los sensores a bordo de los satélites, por lo que en este caso el operador observacional es la función que mapea temperatura

a refractividad. En esta tesis \mathcal{M}_Ω , \mathcal{H} y \mathbf{R} son asumidos conocidos, mientras \mathbf{Q} se asume total o parcialmente desconocida.

Un proceso de Markov escondido cumple, por definición, la siguiente propiedad:

$$p(\mathbf{Y}_0 = \mathbf{y}_0, \dots, \mathbf{Y}_K = \mathbf{y}_K | \mathbf{X}_0 = \mathbf{x}_0, \dots, \mathbf{X}_K = \mathbf{x}_K) = \prod_{k=0}^K p(\mathbf{y}_k | \mathbf{x}_k). \quad (2.3)$$

Además, al ser $\{\mathbf{X}_i\}_{i=1}^K$ un proceso de Markov cumple por definición:

$$p(\mathbf{X}_{n+1} = \mathbf{x}_{n+1} | \mathbf{X}_n = \mathbf{x}_n, \mathbf{X}_{n-1} = \mathbf{x}_{n-1}, \dots, \mathbf{X}_0 = \mathbf{x}_0) = p(\mathbf{X}_{n+1} = \mathbf{x}_{n+1} | \mathbf{X}_n = \mathbf{x}_n) \quad (2.4)$$

para todo $n \in 1, \dots, K-1$. Esta ecuación se denomina propiedad de Markov y establece que el comportamiento (aleatorio) del proceso en tiempo $n+1$ depende únicamente del estado en tiempo n y no de la historia previa del proceso. Generalmente, se dice que el “futuro” depende únicamente del estado “actual” del proceso. De manera similar, el valor de la variable observada \mathbf{y}_k solo depende del valor de la variable oculta \mathbf{x}_k , ambas en el tiempo k .

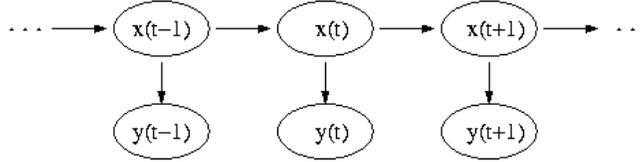


Figura 2.1: Dependencias en un HMM

En resumen, se tiene el siguiente sistema dinámico discreto estocástico:

$$\begin{cases} \mathbf{x}_k = \mathcal{M}_\Omega(\mathbf{x}_{k-1}) + \eta_k, & \eta_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \epsilon_k, & \epsilon_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{cases} \quad (2.5)$$

Donde:

- \mathbf{x}_k , es una realización de la variable aleatoria \mathbf{X}_k , y es el estado del modelo en el tiempo t_k , n -dim.
- η_k es el error del modelo.
- \mathbf{y}_k es la medición, m -dim.
- \mathcal{M}_Ω es el modelo.
- \mathcal{H} es el operador de observación.

2.2. Métodos y conceptos de Asimilación de Datos

2.2.1. Conceptos

Las definiciones de esta tesis van a estar enfocadas a las ciencias de la atmósfera. Los siguientes conceptos, siguiendo las definiciones de Bouttier (2002) [2], abarcan las principales componentes de la asimilación de datos. El objetivo de asimilación de datos es la estimación [20] del estado de la atmósfera, o del sistema en cuestión, combinando estadísticamente dos fuentes de información: por un lado pronósticos o predicciones del estado del sistema, y por otro observaciones indirectas, parciales y ruidosas.

La palabra “predicción” se utilizará exclusivamente para referirse al pronóstico temporal del estado, es decir dado un estado a un tiempo k se predice temporalmente el estado al tiempo $k + 1$ (asumiendo un modelo de Markov conocido).

El producto resultante de la asimilación de datos es denominado análisis. El análisis es un estado estimado del sistema para un tiempo dado. Es representado como un vector de la dimensión del estado del modelo. Un análisis puede ser útil en sí mismo para un diagnóstico de la atmósfera, y puede ser usado como dato de entrada de otra operación. Notablemente puede ser usado como estado inicial de un pronóstico numérico del tiempo, o como recuperador de datos para ser usado como una pseudo-observación.

La información objetiva básica que se puede usar para producir el análisis es un conjunto de valores observados proporcionados por mediciones del sistema en consideración. Para que sea un problema de optimización bien planteado, es necesario contar con alguna fuente de información extra, denominada *background*, que representa una inferencia a priori del estado del modelo. La información de background puede ser climatológica o un estado trivial; puede ser generado por la salida de un análisis previo, usando algún supuesto de consistencia en el tiempo del estado de modelo.

Actualmente, es común que la información de background sea producida por un modelo matemático de predicción que evoluciona al estado del sistema desde el tiempo anterior al tiempo actual. A través de la combinación de estas dos fuentes de información, observaciones y background, a distintos tiempos, se espera que la información se acumule en el estado del modelo y se propague a todas las variables del modelo.

Se entiende por asimilación de datos a todo el conjunto de técnicas que permiten realizar la estimación del estado escondido del sistema a partir de las fuentes de información mencionadas. Dependiendo de la técnica el resultado de la asimilación puede ser una densidad (como en el caso del filtro de partículas), un valor óptimo (como en la asimilación variacional) o los parámetros de una densidad (como en el caso del filtro de Kalman tradicional). El término análisis se utiliza para referirnos al resultado de la asimilación más allá que su definición específica dependiera del método utilizado.

Entre las técnicas de asimilación de datos hay dos enfoques básicos de asimilación: asimilación *secuencial*, que solo considera observaciones hechas en el pasado hasta el tiempo del análisis, el cual es el caso de los sistemas de asimilación en tiempo real, y *no-secuencial*, o asimilación retrospectiva, donde las observaciones del futuro son usadas, por ejemplo, en un ejercicio de reanálisis. Otra distinción que puede hacerse entre los métodos es si son *continuos* o *intermitentes* en el tiempo. En un método intermitente las observaciones pueden ser procesadas en un intervalo pequeño, lo cual es usualmente conveniente para la técnica. En un método continuo, las observaciones son consideradas para periodos más largos, y la corrección del estado de análisis es suave en el tiempo, lo cual es físicamente más real. Los cuatro tipos de asimilación son descritos simbólicamente en la Figura 2.2.

Algunas técnicas de asimilación han sido usadas para meteorología y oceanografía. En la Figura 2.3 se observan algunos métodos, ordenados por complejidad. Entre ellos difieren su costo numérico, su optimalidad, y su idoneidad para asimilación de datos en tiempo real.

Dentro de la metodología más simple se encuentra la conocida por interpolación de observaciones. Luego podemos encontrar un conjunto de métodos llamados métodos primitivos, de los cuales podemos destacar: ajuste polinómico, spline cúbico, correcciones sucesivas (SCM), y método de Cressman. Otro método, un poco más complejo, pero también simple, es el de Interpolación óptima. Se denomina de esa forma o interpolación estadística con estimación de mínimos cuadrados.

En las subsecciones siguientes se da una explicación mínima sobre teoría de estimación, y sobre los métodos 3D-Var, Filtro de Kalman (KF), Filtro de Kalman Extendido (EKF), 4D-Var, Filtro de Kalman por Ensamble (EnKF) y Suavizado de Kalman por Ensamble (EnKS). Estos son los métodos más utilizados en la asimilación de datos.

Por último, nos referimos por sistema de asimilación, al conjunto de implementaciones de un modelo de Markov escondido, incluyendo el modelo de predicción del estado, el operador o modelo observacional, y la técnica de estimación o inferencia del estado escondido.

2.2.2. Teoría de estimación

El objetivo último de la asimilación de datos es el conocimiento de la función de distribución de probabilidad (pdf) posterior, esto es, en el sistema planteado en (2.21), conocer la pdf condicional del estado \mathbf{x}_k dada la secuencia $\mathbf{Y}_l = \{\mathbf{y}_1, \dots, \mathbf{y}_l\}$ de observaciones de dimensión m . [21]

A partir de la relación entre k y l se clasifica el problema de estimación en tres tipos:

- $k < l$ problema de suavizado (reanálisis)
- $k = l$ problema de filtrado (análisis)
- $k > l$ problema de predicción

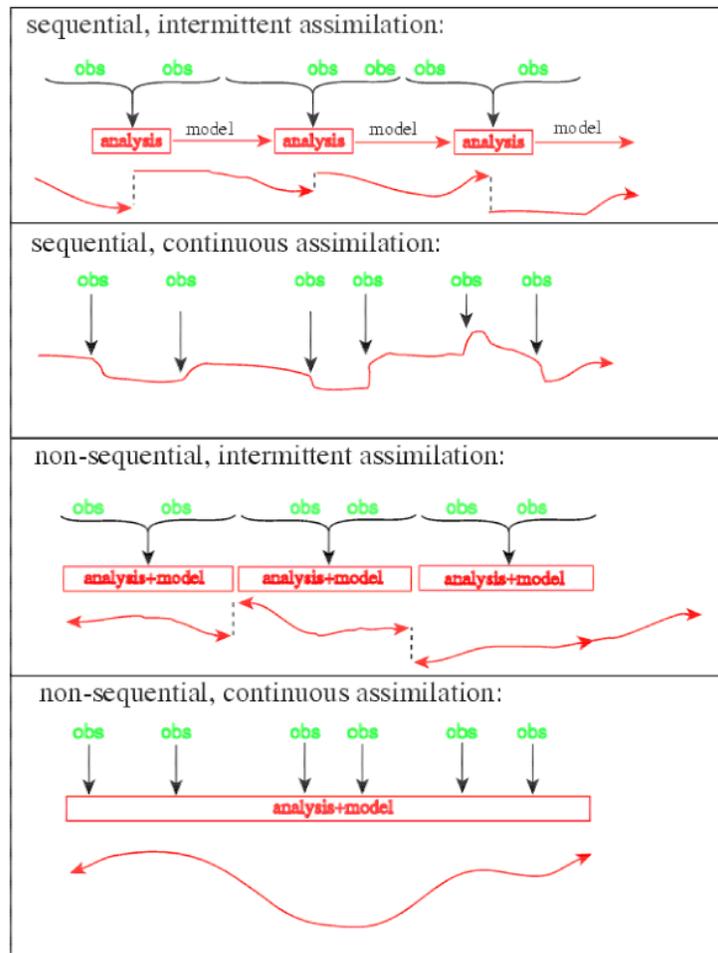


Figura 2.2: [Figura obtenida de Bouttier et al. (1999) [2]] Representación de cuatro estrategias de asimilación de datos, en función del modo en que son procesadas las observaciones en el tiempo para producir una secuencia de estados asimilados (curva por debajo en cada panel).

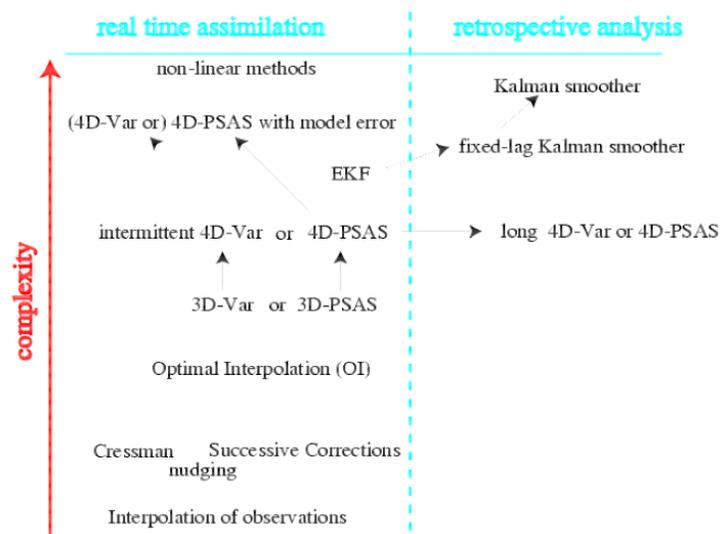


Figura 2.3: [Figura obtenida de Bouttier et al. (1999) [2]] Resumen histórico de los algoritmos de asimilación de datos usados en meteorología y oceanografía, clasificados aproximadamente acorde a su complejidad y costo de implementación, y su aplicabilidad a problemas de tiempo real. Actualmente, los más comúnmente usados para aplicaciones operativas son OI, 3D-Var y 4D-Var.

Dado que encontrar la pdf resulta en general imposible, lo que se propone es determinar los primeros momentos o ciertas propiedades de la pdf posterior:

- Estimador de varianza mínima (MV): se busca determinar la media de la pdf condicional posterior.
- Estimador de máxima a posteriori (MAP): se busca determinar la moda de la pdf condicional.

Si se asume que no se conoce nada de la estadística de los errores (ni sobre las observaciones ni sobre el estado), el problema es totalmente determinístico. Lo que se busca en este caso es minimizar la diferencia entre la observación \mathbf{y} y el estado del modelo proyectado sobre el espacio de las observaciones $\mathcal{H}(\mathbf{x})$. En este caso se trabaja con un estimador por cuadrados mínimos (LS).

2.2.3. Asimilación variacional de 3 dimensiones

La asimilación de datos variacional de 3 dimensiones, o 3D-Var, ha sido una de las técnicas más usadas en la producción de pronósticos meteorológicos. Hoy en día sigue siendo usada, pero muchos se están inclinando más por el 4D-Var o EnKF.

Hipótesis del 3D-Var:

- De las observaciones del estado \mathbf{x} , $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \epsilon$, se asume que ϵ es una variable aleatoria normal con media 0 y matriz de covarianza \mathbf{R} , y por lo tanto, $\mathbf{y}|\mathbf{x}$ es una variable aleatoria con media $\mathcal{H}(\mathbf{x})$ y covarianza \mathbf{R} .

$$\epsilon \sim \mathcal{N}(0, \mathbf{R})$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathcal{H}(\mathbf{x}), \mathbf{R})$$

- La distribución del estado se asume normal, con media \mathbf{x}^b y matriz de covarianza \mathbf{B} .

$$\mathbf{x}_k \sim \mathcal{N}(\mathbf{x}^b, \mathbf{B})$$

La idea del 3D-Var es estimar el máximo de la probabilidad condicional a priori $p_{\mathbf{x}|\mathbf{y}}$. Usando la regla de Bayes y las hipótesis de gaussianidad se tiene que:

$$p_{\mathbf{x}|\mathbf{y}} = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y})} \quad (2.6)$$

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \left[(\mathbf{x} - \mathbf{x}^b)^T (\mathbf{B})^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})) \right] \right\} \quad (2.7)$$

Se desentiende de la $p_{\mathbf{y}}(\mathbf{y})$ del denominador porque no influye en la estimación, ya que se quiere determinar \mathbf{x} . Luego, lo que se quiere encontrar es el

máximo argumento de la exponencial, i.e. la moda, lo que equivale al mínimo de la función de costo $J(\mathbf{x})$ definida por:

$$J(\mathbf{x}) = -\log p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}), \text{ que equivale a:}$$

$$J(\mathbf{x}) = \frac{1}{2} \left[(\mathbf{x} - \mathbf{x}^b)^T (\mathbf{B})^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})) \right]. \quad (2.8)$$

Entonces 3D-Var es una estimación de Probabilidad Máxima a Posteriori donde se asume estadística gaussiana. Notar que en ningún momento se asume que \mathcal{H} sea lineal. Si \mathcal{H} es lineal y se asume gaussianidad, 3D-Var y OI (interpolación óptima) coinciden.

2.2.4. Filtro de Kalman

El filtro de Kalman, o KF, es una técnica de asimilación de datos, que como tal, dada una estimación a priori (background) $\hat{\mathbf{x}}_k^f$, que es un pronóstico del sistema para el tiempo t_k , y un conjunto de mediciones \mathbf{y}_k , pretende estimar un análisis $\hat{\mathbf{x}}_k^a$. Esta técnica asume el modelo \mathcal{M}_Ω y el operador de observaciones \mathcal{H} lineales [21]. Este último se representa con la matriz \mathbf{H} de dimensión $m \times n$. El análisis para el tiempo t_k se calcula de la siguiente manera:

$$\hat{\mathbf{x}}_k^a = \hat{\mathbf{x}}_k^f + \tilde{\mathbf{K}}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^f \right) \quad (2.9)$$

Se debe determinar el peso $\tilde{\mathbf{K}}_k$ de la innovación, de tal forma que minimice la varianza del error del análisis. Si esta está definida por:

$$\begin{aligned} J_k^a &= \left\langle \mathbf{e}_k^a \mathbf{S}_k (\mathbf{e}_k^a)^T \right\rangle \\ &= \left\langle \text{Tr} \left(\mathbf{S}_k \mathbf{e}_k^a (\mathbf{e}_k^a)^T \right) \right\rangle \\ &= \text{Tr} (\mathbf{S}_k \mathbf{P}_k^a) \end{aligned} \quad (2.10)$$

donde \mathbf{S}_k es una matriz de scaling definida positiva, lo que se busca entonces es determinar el $\tilde{\mathbf{K}}_k$ que minimice J_k^a . Derivando e igualando a cero J_k^a se obtiene la matriz de peso óptima \mathbf{K}_k . [4]

$$\begin{aligned} \frac{dJ}{d\tilde{\mathbf{K}}_k} &= \mathbf{S}_k \left[-2\mathbf{H}_k \mathbf{P}_k^f \left(\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k \right)^T + 2\mathbf{R}_k \tilde{\mathbf{K}}_k^T \right] \\ \tilde{\mathbf{K}}_k = \mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1} \end{aligned} \quad (2.11)$$

El estimador del análisis por medio del filtro de Kalman es entonces un estimador de mínima varianza.

Para $\tilde{\mathbf{K}}_k$ óptimo se tiene que:

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f. \quad (2.12)$$

Por lo cual, la covarianza del pronóstico puede evolucionar en el tiempo, a diferencia de OI y 3D-Var que es estática, ya que son esquemas independientes del tiempo.

Filtro de Kalman extendido

El filtro de Kalman está definido para un modelo lineal y un operador de observación lineal. Para extenderlo a casos no lineales se aproxima con Taylor ambas funciones. En esto se basa el filtro de Kalman extendido (Extended Kalman filter, o EKF), para luego utilizar las fórmulas del filtro de Kalman, usando las tangentes lineales (detalles de las hipótesis en [2]).

2.2.5. Asimilación variacional de 4 dimensiones

En la asimilación de datos variacional de 4 dimensiones, o 4D-Var, se piensa un problema donde las observaciones no están distribuidas solo espacialmente, sino también temporalmente. Se busca la condición inicial, a partir de la cual, se realice la trayectoria del modelo más cercana posible a las observaciones. A esta condición inicial, se la denomina condición inicial óptima. Se observa en la Figura 2.4 una representación gráfica de esto.

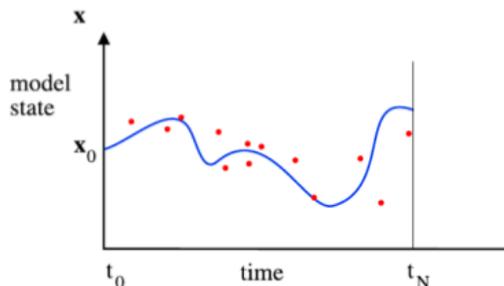


Figura 2.4: 4D-Var: se busca la condición inicial, a partir de la cual, se realice la trayectoria del modelo más cercana posible a las observaciones. REF

En fórmulas, esto sería:

$$J(\mathbf{x}_0) = \frac{1}{2} \int_0^T \langle (\mathbf{x}(t, \mathbf{x}_0) - \mathbf{y}(t)) | (\mathbf{x}(t, \mathbf{x}_0) - \mathbf{y}(t)) \rangle dt \quad (2.13)$$

$\mathbf{y}(t)$ son las observaciones continuas en el tiempo, realizadas en los mismos puntos del estado. $\langle | \rangle$ representa un producto interno en algún espacio. La función de costo a minimizar es una función de las condiciones iniciales, los parámetros del modelo y las condiciones de contorno. Discretizando el espacio y

tiempo, se tiene:

$$J(\mathbf{x}_0) = \frac{1}{2} [\mathbf{x}_0 - \mathbf{x}_0^f]^T (\mathbf{P}^f)^{-1} [\mathbf{x}_0 - \mathbf{x}_0^f] + \frac{1}{2} \sum_i [\mathbf{y}_i - \mathbf{H}(\mathbf{x}_i)]^T \mathbf{R}^{-1} [\mathbf{y}_i - \mathbf{H}(\mathbf{x}_i)] \quad (2.14)$$

La evolución del modelo viene dada por $\mathbf{x}_i = \mathcal{M}_\Omega(\mathbf{x}_{i-1})$, donde \mathbf{x}_0 es la condición inicial y el modelo es no lineal. Para calcular el gradiente de J desde $k=0$ a $k=N$ se utiliza el tangente lineal del modelo evaluado en \mathbf{x}_k y a base de este se calcula el modelo adjunto \mathbf{M}_k y se lo denota \mathbf{M}_k^* .

$$\nabla_{\mathbf{x}_0} J = \sum_{k=0}^N \mathbf{M}_0^* \cdots \mathbf{M}_{k-1}^* \nabla_{\mathbf{x}_k} J \quad (2.15)$$

El gradiente es necesario para minimizar la función de costo de manera iterativa. Por lo tanto, para cada iteración, se requiere evaluar a $J(\mathbf{x}_0)$ y a $\nabla_{\mathbf{x}_0} J$, es decir, se requiere una integración del modelo (no lineal) hacia adelante y luego una integración del adjunto.

2.2.6. Filtro de Kalman por ensambles

En el filtro de Kalman por ensambles (EnKF) se asume gaussianidad y se aproxima la media y la covarianza de las distribuciones a partir de un muestreo de Montecarlo [22][23]. En general se puede representar una densidad de probabilidad por la densidad empírica de un conjunto grande de muestras aleatorias (un ensamble). Esta es la base de los métodos de Monte Carlo secuenciales.

En Doucet et al. 2001 [24] y Del Moral et al. 2006 [25] se denomina partículas o miembros de un ensamble al conjunto de N_e muestras aleatorias $\{x_{k,j}\}$ $j = 1, \dots, N_e$ que en general pueden estar pesadas con pesos $w_{k,j}$. Se demuestra que esta distribución empírica converge asintóticamente a la distribución $p(\mathbf{x}_k)$.

Por ende, a partir de estas, se puede aproximar la esperanza del estado real, o de una función para cualquier función determinada para el estado, de la siguiente manera:

$$\mathcal{E}(f(\mathbf{x})) = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \approx \sum_{j=1}^{N_e} w_{k,j} f(\mathbf{x}_j) \quad (2.16)$$

lo cual se denomina integración de Montecarlo. En particular, si todas las partículas tienen el mismo peso, se puede aproximar la media del estado real por:

$$\hat{\mathbf{x}}_k = \frac{1}{N_e} \sum_j^{N_e} \mathbf{x}_{k,j}, \quad (2.17)$$

y la covarianza por:

$$\hat{\mathbf{P}}_k = \frac{1}{N_e} \sum_j^{N_e} (\mathbf{x}_{k,j} - \hat{\mathbf{x}}_k) (\mathbf{x}_{k,j} - \hat{\mathbf{x}}_k)^\top \quad (2.18)$$

conocida como covarianza muestral, o sample covariance en inglés. Se tiene así una manera alternativa de calcular las covarianzas del pronóstico y del análisis.

Supongamos que se tiene un ensamble $\mathbf{x}_{0,1:N_e}$ que son N_e muestras de la densidad de probabilidad a priori del estado inicial $p(\mathbf{x}_0)$, el estado en el tiempo 0, se obtiene un ensamble de muestra del estado $p(\mathbf{x}_1)$ (densidad de la predicción) haciendo:

$$\mathbf{x}_{1,j} = \mathcal{M}_\Omega(\mathbf{x}_{0,j}) + \eta_{1,j}$$

donde $\eta_{1,j}$ son realizaciones del error de modelo, con $\eta_{1,j} \sim \mathcal{N}(0, \mathbf{Q})$. En la Figura 2.5 se representa el proceso de asimilación en un tiempo t utilizando EnKF.

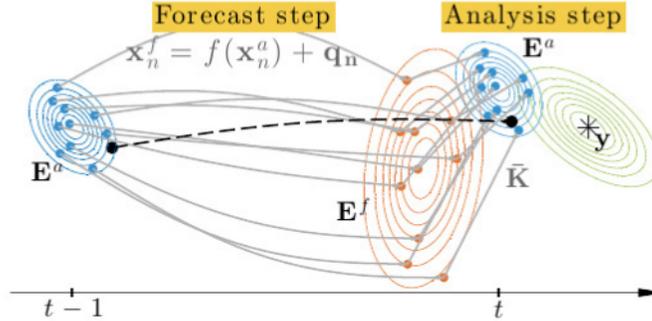


Figura 2.5: EnKF: Representación de un ensamble de análisis (E^a) en un tiempo $t - 1$, el avance de este por medio del modelo f y realizaciones q_n del error de modelo (análogo al $\eta_{t,n}$), generando el ensamble de pronóstico en el tiempo t (E^f), y el nuevo ensamble de análisis (E^a) en este tiempo luego de asimilar la observación y . (Fig. obtenida de [26])

El ensamble de análisis se obtiene luego de asimilar cada miembro del ensamble del pronóstico como en un filtro de Kalman. Se obtiene el miembro j del ensamble de análisis, $\mathbf{x}_{k,j}^a$, de la siguiente manera:

$$\mathbf{x}_{k,j}^a = \mathbf{x}_{k,j}^f + \mathbf{K}_k \left(\mathbf{y}_k - \mathcal{H}(\mathbf{x}_{k,j}^f) - \epsilon_{k,j} \right) \quad (2.19)$$

donde la matriz de peso viene dada por:

$$\hat{\mathbf{K}}_k = \hat{\mathbf{P}}_k^f \mathbf{H}^\top \left(\mathbf{H} \hat{\mathbf{P}}_k^f \mathbf{H}^\top + \mathbf{R}_k \right)^{-1}. \quad (2.20)$$

En (2.20) se usa la covarianza muestral del pronóstico, y $\epsilon_{k,j}$ en (2.19) es una realización del error de las observaciones (observaciones perturbadas) [27].

Burgers et al. (1996/8) [22] son los que hacen la metodología. Originalmente, la propuesta de Evensen (1994) [5] no incluía el término de error de las observaciones y esto resulta en un análisis inconsistente con una covarianza errónea. Posteriormente, esto fue notado por Burgers incluyendo el término de error de las observaciones (realizaciones aleatorias) y logrando consistencia de esta manera en el ensamble con la covarianza del análisis prevista.

En resumen, se tiene el sistema dinámico discreto estocástico:

$$\begin{cases} \mathbf{x}_k = \mathcal{M}_\Omega(\mathbf{x}_{k-1}) + \eta_k, & \eta_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \epsilon_k, & \epsilon_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{cases} \quad (2.21)$$

Donde:

- \mathbf{x}_k es el estado del modelo en el tiempo t_k , n -dim.
- η_k es el error del modelo
- \mathbf{y}_k m -vector es la medición.
- \mathcal{M}_Ω es el modelo
- \mathcal{H} es el operador de observación

En el cual se desea estimar \mathbf{x}_k . Con EnKF, las distribuciones del pronóstico y el análisis son aproximadas por muestras discretas $\mathbf{x}_{k,j}^f$ y $\mathbf{x}_{k,j}^a$ respectivamente para cada tiempo t_k con $k = 1, \dots, K$, donde $j = 1, \dots, N_e$ y N_e es el número de miembros del ensamble. La muestra es inicializada con $\mathbf{x}_{0,j}^a \sim \mathcal{N}(0, B)$. Luego, se realizan alternadamente los pasos de pronóstico y de análisis para $k = 1, \dots, K$ [18][22].

- Paso de pronóstico EnKF: El paso de pronóstico propaga el ensamble de análisis desde el tiempo t_{k-1} al tiempo t_k , y computa la muestra de observaciones pronosticadas, usando las ecuaciones de (2.21):

$$\begin{aligned} \mathbf{x}_{k,j}^f &= f_{k-1,k} \left(\mathbf{x}_{k-1,j}^a \right) + \eta_{k,j}, & j = 1, \dots, N_e \\ \mathbf{y}_{k,j}^f &= \mathbf{H} \mathbf{x}_{k,j}^f + \epsilon_{k,j}, & j = 1, \dots, N_e \end{aligned} \quad (2.22)$$

con $\eta_{k,j} \sim \mathcal{N}(0, \mathbf{Q})$, $\epsilon_{k,j} \sim \mathcal{N}(0, \mathbf{R})$, $\mathbf{y}_{k,j}^f$ la observación pronosticada para la muestra j , y f el modelo de pronóstico temporal del estado, es decir la función que dado un estado en un tiempo k , predice temporalmente el estado para el tiempo $k+1$.

- Paso de análisis EnKF: El ensamble de pronóstico $\mathbf{x}_{k,j}^f$ es corregido con las nuevas observaciones \mathbf{y}_k para obtener el análisis $\mathbf{x}_{k,j}^a$:

$$\begin{aligned} \mathbf{K}_k^f &= \mathbf{P}_k^f \mathbf{H}^T \left[\mathbf{H} \mathbf{P}_k^f \mathbf{H}^T + \mathbf{R} \right]^{-1} \\ \mathbf{x}_{k,j}^a &= \mathbf{x}_{k,j}^f + \mathbf{K}_k^f \left(\mathbf{y}_k - \mathbf{y}_{k,j}^f \right), j = 1, \dots, N_e \end{aligned} \quad (2.23)$$

Con \mathbf{P}_k^f la covarianza muestral (o sample covariance) de la muestra $\mathbf{x}_{k,j}^f$, $j = 1, \dots, N_e$.

El modelo f con el que se asimila puede o no coincidir con el modelo \mathcal{M}_Ω con el que se generó la verdad. Si coinciden se dice que el modelo f es perfecto.

2.2.7. Suavizado de Kalman por ensambles

Los métodos descritos (4D-Var, KF, EnKF) permiten resolver el problema de estimación de filtrado (análisis), en el cual, según lo planteado en 2.2.2, se desea estimar la función de densidad de probabilidad (pdf) posterior del estado \mathbf{x}_k dado una secuencia de observaciones $\mathbf{y}_{1:t}$, $p(\mathbf{x}_k|\mathbf{y}_{1:t})$, para el caso particular donde $k = t$.

Cualquier planteo, con $k \leq t$, distinto al anterior, es un problema de suavizado. Por ejemplo, el de estimar la probabilidad conjunta $p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k})$. También lo es el de resolver el problema marginal de estimar $p(\mathbf{x}_k|\mathbf{y}_{1:t})$ para el caso que se tengan observaciones tanto pasadas, como presente, y futuras, es decir, cuando $k < t$. Varios algoritmos de suavizado basados en el Filtro de Kalman han sido presentados y usados en aplicaciones de geofísica, y se denomina Suavizado de Kalman (Kalman Smoother, KS). En Cosme (2012) [28] se plantea cinco enfoques distintos: tres para el problema del suavizado conjunto, y dos para el suavizado marginal. Nos enfocamos en este último, que es el que se utilizará posteriormente en el desarrollo de la técnica basada en el algoritmo EM.

Para el problema del suavizado marginal se plantea *Forward-backward smoother* (suavizado adelante-atrás), también llamado suavizado Rauch–Tung–Striebel (RTS), y *Two-filter smoother* (suavizado dos filtros). El primero es el que más se ha utilizado en aplicaciones geofísicas.

Se presentan a continuación las ecuaciones para el caso lineal gaussiano, para resolver el problema de estimar la media y covarianza de $p(\mathbf{x}_k|\mathbf{y}_{1:K})$, donde $k \leq K$, por medio de $\mathbf{x}_{k|1:K}^s$ y $\mathbf{P}_{k|1:K}^s$.

$$\begin{aligned} \mathbf{K}_k^s &= \mathbf{P}_{k|1:k}^a \mathbf{M}_{k,k+1}^T \left(\mathbf{P}_{k+1|1:k}^f \right)^{-1} \\ \mathbf{x}_{k|1:K}^s &= \mathbf{x}_{k|1:k}^a + \mathbf{K}_k^s \left(\mathbf{x}_{k+1|1:K}^s - \mathbf{x}_{k+1|1:k}^f \right) \\ \mathbf{P}_{k|1:K}^s &= \mathbf{P}_{k|1:k}^a - \mathbf{K}_k^s \left(\mathbf{P}_{k+1|1:k}^f - \mathbf{P}_{k+1|1:k}^s \right) \mathbf{K}_k^{sT} \end{aligned} \quad (2.24)$$

Estas ecuaciones (2.24) son aplicadas después de haber pasado un filtro de Kalman sobre el intervalo determinado por el conjunto de observaciones. Esto se realiza iterativamente, con k que corre desde $K - 1$ hasta 1, y se inicializa tomando $\mathbf{x}_k^s = \mathbf{x}_k^a$ y $\mathbf{P}_k^s = \mathbf{P}_k^a$. Esta idea puede extenderse al caso no-lineal utilizando ensambles, y se denomina Suavizado de Kalman por ensambles (Ensemble Kalman Smoother, EnKS).

2.2.8. Parámetroización de la Covarianza del pronóstico

En la práctica el filtro de Kalman por ensamble puede fallar en la sincronización con la verdad. Incluso cuando el modelo es perfecto (sin error) el filtro

tiende a subestimar la incertidumbre [29] de la estimación del estado. Independientemente de la causa, subestimar la incertidumbre conduce a un exceso de seguridad en la estimación del estado de background, y, por lo tanto, las observaciones pierden peso en el cálculo del análisis. Si la discrepancia se vuelve demasiado grande sobre el tiempo, las observaciones son esencialmente ignoradas por el análisis, y la dinámica de la asimilación de datos empieza a desacoplarse de la verdad.

Una manera de contrarrestar esta tendencia y dejar de subestimar la incertidumbre en la estimación del estado, es incrementando el error del pronóstico por medio de parámetros ajustables. Este es un procedimiento *ad hoc*. De estos existen los métodos de la inflación aditiva y multiplicativa.

En EnKF, el error de pronóstico involucrado en el ciclo de asimilación k proviene de dos fuentes: la matriz de error de modelo \mathbf{Q}_k , y la matriz de covarianza del ensamble del pronóstico \mathbf{P}_k^f . Por lo que la inflación aditiva también es congruente con el modelo de Markov escondido con error de modelo.

Para determinar el error de modelo óptimo del modelo de Markov escondido (sec. (2.1)-(2.2)) proponemos re-escribir la matriz de covarianza del error de modelo por $\beta\mathbf{Q}$. Donde β es un factor multiplicativo, parámetro a estimar, y $\beta\mathbf{Q}$ es una matriz de covarianza del error de modelo asumida constante y representada a través del error de modelo climatológico.

Por otro lado, la *inflación multiplicativa* propone separar los miembros del ensamble de pronóstico, aumentando su perturbación con respecto a la media, de tal forma que la nueva matriz de covarianza pase a ser $\alpha\mathbf{P}_k^f$, con α mayor a 1, como se muestra en la ecuación (2.25). Esto se realiza antes de asimilar la observación \mathbf{y}_k correspondiente del tiempo k .

$$\begin{aligned}\bar{\mathbf{x}}_k^f &\leftarrow \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{x}_{k,j}^f \\ \mathbf{x}_{k,j}^f &\leftarrow \bar{\mathbf{x}}_k^f + \sqrt{\alpha} \left(\mathbf{x}_{k,j}^f - \bar{\mathbf{x}}_k^f \right)\end{aligned}\quad (2.25)$$

Si se incorpora error de modelo, y se lo amplifica con un parámetro β , y, se amplifica la covarianza del ensamble con un parámetro α , la matriz de covarianza de la predicción efectiva que proponemos trabajar en esta tesina viene entonces definida por $\mathbf{P}_k^f = \alpha\hat{\mathbf{P}}_k^f + \beta\mathbf{Q}$ donde $\hat{\mathbf{P}}_k^f$ es la covarianza muestral del ensamble de las predicciones.

2.3. Estimación de máxima verosimilitud aproximada

2.3.1. Aproximación de la función de log-verosimilitud

Esta subsección se basa en el trabajo de Pulido et al. (2017) [16] en el cual se presenta una aproximación a la función de log-verosimilitud, denominada log-verosimilitud de innovación o innovation log-likelihood en inglés. También nos referiremos a ella más adelante como log-verosimilitud aproximada.

Se descompone la función de densidad de probabilidad de las observaciones en el siguiente producto

$$p(\mathbf{y}_{1:K}; \theta) = \prod_{k=1:K} p(\mathbf{y}_k | \mathbf{y}_{1:k-1}; \theta), \quad (2.26)$$

donde θ representa los parámetros estadísticos, y con la convención de que $y_{1:0} = \emptyset$. Tomando logaritmo y reescribiendo cada término en función de la integral sobre el estado x_k , se tiene

$$l(\theta) = \sum_{k=1}^K \ln p(\mathbf{y}_k | \mathbf{y}_{1:K}; \theta) = \sum_{k=1}^K \ln \int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}; \theta) d\mathbf{x}_k. \quad (2.27)$$

Si se asume Gaussianidad, y lineales tanto el modelo como el operador de observaciones, la integral de la ecuación (2.27) es exactamente la distribución de análisis dada por el filtro de Kalman. La verosimilitud de las observaciones condicionadas por el estado en cada tiempo está dada por

$$p(\mathbf{y}_k | \mathbf{x}_k) = \left[(2\pi)^{M/2} |\mathbf{R}|^{1/2} \right]^{-1} \exp \left[-\frac{1}{2} (\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k))^T \mathbf{R}^{-1} (\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k)) \right], \quad (2.28)$$

y la distribución a priori del pronóstico está dada por

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}; \theta) = \left[(2\pi)^{N/2} |\mathbf{P}_k^f|^{1/2} \right]^{-1} \exp \left[-\frac{1}{2} (\mathbf{x}_k - \mathbf{x}_k^f)^T (\mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \mathbf{x}_k^f) \right] \quad (2.29)$$

donde $\mathbf{x}_k^f = \mathcal{M}_\Omega(\mathbf{x}_{k-1}^a) + \eta_k$ es el pronóstico, \mathbf{x}_{k-1}^a es la estimación de la media del análisis en el tiempo $k-1$, y \mathbf{P}_k^f es la matriz de covarianza del pronóstico del filtro[30]. Notar que la ecuación (2.29) se cumple solo en forma aproximada para modelos dinámicos no lineales.

Asumiendo linealidad del operador \mathbf{H} Reemplazando (2.28) y (2.29) en (2.27) y calculando la integral sobre \mathbf{x}_k , se obtiene

$$l(\theta) \approx -\frac{1}{2} \sum_{k=1}^K \left[(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k^f)^T (\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k) + \ln(|\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T + \mathbf{R}|) \right] + C \quad (2.30)$$

donde C es una constante independiente de θ .

2.3.2. Algoritmo Nelder-Mead

Este algoritmo es un método numérico para minimizar una función objetivo en un espacio multidimensional, y optimiza basado solo en los valores de la

función objetivo. Tiene la ventaja de no necesitar gradientes para operar, aunque esto hace que sea más lento y, por lo tanto, más ineficiente.

El método trabaja con un politipo de $N+1$ vértices en donde N es el número de parámetros, o dimensión del estado donde se optimiza [31]. Por ejemplo, si se optimiza en un espacio de dimensión dos porque hay dos parámetros a optimizar, el método trabajará con un politipos de tres vértices, es decir, con triángulos.

El método se basa en los siguientes pasos:

1. Se calcula las evaluaciones funcionales en los vértices del politipo, y se ordenan en función de los valores de dichas evaluaciones.

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$$

Se chequea criterio de corte.

2. Se calcula x_0 , el centroide de todos los puntos excepto de x_{n+1} .
3. Reflexión: Se computa el punto reflejado $x_r = x_0 + k(x_0 - x_{n+1})$, con $k > 0$. Si ocurre que $f(x_1) \leq f(x_r) < f(x_n)$, se obtiene un nuevo politipo de reemplazar el peor punto, x_{n+1} por x_r .
4. Expansión: Si ocurre que $f(x_r) < f(x_1)$, se computa el punto expandido $x_e = x_0 + t(x_r - x_0)$, con $t > 0$. Si el punto expandido es mejor que el reflejado, es decir, $f(x_e) < f(x_r)$, se obtiene un nuevo politipo reemplazando el peor punto, x_{n+1} por x_e , y se vuelve al paso 1. Si no, se reemplaza x_{n+1} por x_r y se vuelve al paso 1.
5. Contracción: Si $f(x_r) > f(x_n)$, se computa $x_c = x_0 + r(x_{n+1} - x_0)$, con $0 < r < 0,5$. Si $f(x_c) < f(x_{n+1})$, el nuevo politipo se obtiene de reemplazar x_{n+1} por x_c , y se vuelve al paso 1.
6. Encogimiento: Se reemplaza todos los puntos, excepto el mejor, x_1 , por $x_i = x_1 + s(x_i - x_1)$ y vuelve al paso 1.

Los valores k, t, r y s son los coeficientes de reflexión, expansión, contracción y encogimiento respectivamente, cuyos valores estándares son: $k = 1$, $t = 2$, $r = 1/2$ y $s = 1/2$.

Para la reflexión, ya que x_{n+1} es el vértice con el valor más grande asociado entre los vértices, se puede esperar encontrar un valor más bajo en la reflexión de x_{n+1} en la cara opuesta formada por todos los vértices x_i excepto x_{n+1} .

Para la expansión, si el punto reflejado x_r es el nuevo mínimo entre los vértices, se puede esperar encontrar valores interesantes a lo largo de la dirección de x_0 a x_r .

Con respecto a la contracción, si $f(x_r) > f(x_n)$, se puede esperar que un valor mejor este dentro del simplex formado por todos los vértices x_i .

Finalmente, el encogimiento es un caso raro en el que en la contracción fuera del punto más grande se incrementa la función, algo que no puede ocurrir lo suficientemente cerca de un mínimo. En ese caso, se contrae hacia el punto más bajo con la expectativa de encontrar un panorama más simple.

Como criterio de parada Nelder y Mead utilizaron la desviación estándar muestral de los valores de función del simplex actual. Si estos caen por debajo de cierta tolerancia, entonces el ciclo se detiene y el punto más bajo en el simplex regresa como el óptimo determinado.

2.4. Algoritmo EM para estimación de parámetros

2.4.1. Problema de estimación

Dada una cadena de Markov escondida y un conjunto de vectores de observaciones distribuidas en el tiempo, $\{\mathbf{y}_k, k = 1, \dots, K\}$ Sea $\theta = \{\mathbf{x}^b, \mathbf{B}, \mathbf{R}, \mathbf{Q}\}$ de una cadena de Markov escondida, donde \mathbf{x}^b y \mathbf{B} es la media y covarianza del estado de inicial del proceso \mathbf{x}_0 , \mathbf{R} es la covarianza del error de medición de las observaciones, y \mathbf{Q} es la covarianza del error del modelo. [18]

$$\begin{cases} \mathbf{x}_k = \mathcal{M}_\Omega(\mathbf{x}_{k-1}) + \eta_k & \eta_k \sim \mathcal{N}(0, \mathbf{Q}) \\ \mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \epsilon_k & \epsilon_k \sim \mathcal{N}(0, \mathbf{R}) \\ & \mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}^b, \mathbf{B}) \end{cases} \quad (2.31)$$

Hiperparámetros: $\theta = \{\mathbf{x}^b, \mathbf{B}, \mathbf{R}, \mathbf{Q}\}$

Buscamos maximizar $p_\theta(\mathbf{y}_{1:K})$. O sea, buscamos

$$\operatorname{argmax}_\theta p_\theta(\mathbf{y}_{1:K}) = \operatorname{argmax}_\theta \ln(p_\theta(\mathbf{y}_{1:K}))$$

Se piensa en la probabilidad de tener las observaciones $y_{1:K}$ como la probabilidad marginal, de la conjunta dada por las observaciones y el vector de estados.

$$p_\theta(y_{1:K}) = \int \cdots \int p_\theta(\mathbf{y}_{1:K}, \mathbf{x}_{0:K}) d\mathbf{x}_0 \cdots d\mathbf{x}_K$$

Paso E: Computar la función auxiliar:

$$G_{\theta^{(i)}}(\theta) = \mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})}(\ln p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K})) \quad (2.32)$$

Paso M: Maximizar $G_{\theta^{(i)}}(\theta)$ sobre $\theta \in \Theta$ y actualizar $\theta^{(i)}$. O sea,

$$G_{\theta^{(i)}}(\theta^{(i+1)}) \geq G_{\theta^{(i)}}(\theta) \forall \theta \in \Theta \quad (2.33)$$

Está demostrado que la actualización de θ incrementa la verosimilitud $p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K})$.

2.4.2. Derivacion del algoritmo para el caso Gaussiano

Paso E

Para el caso de un modelo de Markov escondido con errores Gaussianos es posible determinar la expresión de la función auxiliar. Se necesita computar la log-verosimilitud $\ln p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K})$, para luego calcular la esperanza condicional dada en (2.32) por la función $G_{\theta(i)}$.

$$p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K}) = p_\theta(\mathbf{x}_{0:K})p_\theta(\mathbf{y}_{1:K}|\mathbf{x}_{0:K}) \quad (2.34)$$

Además, por la propiedad de Markov (2.4), se tiene

$$p_\theta(\mathbf{x}_{0:K}) = p_\theta(\mathbf{x}_K|\mathbf{x}_{0:K-1})p_\theta(\mathbf{x}_{0:(K-1)}) = p_\theta(\mathbf{x}_K|\mathbf{x}_{K-1})p_\theta(\mathbf{x}_{0:K-1})$$

Si se sigue desarrollando, se obtiene:

$$p_\theta(\mathbf{x}_{0:K}) = p_\theta(\mathbf{x}_0)p_\theta(\mathbf{x}_1|\mathbf{x}_0) \cdots p_\theta(\mathbf{x}_K|\mathbf{x}_{K-1}) \quad (2.35)$$

Si tomamos observaciones independientes entre sí y que solo dependan del estado actual, tenemos

$$p_\theta(\mathbf{y}_{1:K}|\mathbf{x}_{0:K}) = p_\theta(\mathbf{y}_1|\mathbf{x}_1), \cdots, p_\theta(\mathbf{y}_K|\mathbf{x}_K) \quad (2.36)$$

Reemplazando (2.35) y (2.36) en (2.34), se obtiene:

$$p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K}) = p_\theta(\mathbf{x}_0) \prod_{i=1}^K p_\theta(\mathbf{x}_i|\mathbf{x}_{i-1})p_\theta(\mathbf{y}_i|\mathbf{x}_i)$$

y,

$$\ln p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K}) = \ln p_\theta(\mathbf{x}_0) + \sum_{i=1}^K \ln p_\theta(\mathbf{x}_i|\mathbf{x}_{i-1}) + \sum_{i=1}^K \ln p_\theta(\mathbf{y}_i|\mathbf{x}_i)$$

Incorporando las hipótesis de Gaussianidad dadas en (2.31) podemos calcular estas probabilidades, y tenemos que:

$$\begin{aligned} L(\theta) = \ln p_\theta(\mathbf{x}_{0:K}, \mathbf{y}_{1:K}) &= -\frac{1}{2} \ln |\mathbf{B}| - \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b) \\ &\quad - \frac{K}{2} \ln |\mathbf{Q}| - \frac{1}{2} \sum_{i=1}^K (\mathbf{x}_i - f_{i-1,i}(\mathbf{x}_{i-1}))^T \mathbf{Q}^{-1} (\mathbf{x}_i - f_{i-1,i}(\mathbf{x}_{i-1})) \\ &\quad - \frac{K}{2} \ln |\mathbf{R}| - \frac{1}{2} \sum_{i=1}^K (\mathbf{y}_i - \mathbf{H}\mathbf{x}_i)^T \mathbf{R}^{-1} (\mathbf{y}_i - \mathbf{H}\mathbf{x}_i) + C \end{aligned}$$

C constante que no depende de θ . A lo obtenido se le calcula la $\mathcal{E}_{p_{\theta(i)}(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})} = \mathcal{E}_i$ para computar la función auxiliar (2.32). En particular:

$$\mathcal{E}_{p_{\theta(i)}(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})}((\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b)) = \text{tr}\{\mathbf{B}^{-1} \mathcal{E}_i[(\mathbf{x}_0 - \mathbf{x}^b)(\mathbf{x}_0 - \mathbf{x}^b)^T]\} \quad (2.37)$$

Usando que $\mathcal{E}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}\mathcal{E}(\mathbf{x}\mathbf{x}'))$. Del mismo modo se calcula el resto de los términos y se obtiene:

$$\begin{aligned}
G_{\theta^{(i)}}(\theta) &= \mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_{0:K}|\mathbf{y}_{1:K})}(\ln p_{\theta}(\mathbf{x}_{0:K}, \mathbf{y}_{1:K})) = \mathcal{E}_i L(\theta) \\
&= -\frac{1}{2} \ln|\mathbf{B}| - \frac{1}{2} \text{tr}\{\mathbf{B}^{-1} \mathcal{E}_i[(\mathbf{x}_0 - \mathbf{x}^b)(\mathbf{x}_0 - \mathbf{x}^b)^T]\} \\
&\quad - \frac{K}{2} \ln|\mathbf{Q}| - \frac{1}{2} \text{tr}\{\mathbf{Q}^{-1} \sum_{j=1}^K \mathcal{E}_i[(\mathbf{x}_j - f_{j-1,j} \mathbf{x}_{j-1})(\mathbf{x}_j - f_{j-1,j} \mathbf{x}_{j-1})^T]\} \\
&\quad - \frac{K}{2} \ln|\mathbf{R}| - \frac{1}{2} \text{tr}\{\mathbf{R}^{-1} \sum_{j=1}^K \mathcal{E}_i[(y_j - \mathbf{H}\mathbf{x}_j)(y_j - \mathbf{H}\mathbf{x}_j)^T]\}
\end{aligned}$$

Definiendo las matrices de covarianza muestrales para reducir la ecuación,

$$\begin{aligned}
G_{\theta^{(i)}}(\theta) &= -\frac{1}{2} \ln|\mathbf{B}| - \frac{1}{2} \text{tr}\{\mathbf{B}^{-1} \Sigma_{(0)}^{(i)}\} \\
&\quad - \frac{K}{2} \ln|\mathbf{Q}| - \frac{1}{2} \text{tr}\{\mathbf{Q}^{-1} \sum_{j=1}^K \Sigma_{(j)}^{(i)}\} \\
&\quad - \frac{K}{2} \ln|\mathbf{R}| - \frac{1}{2} \text{tr}\{\mathbf{R}^{-1} \sum_{j=1}^K \Omega_{(j)}^{(i)}\}
\end{aligned} \tag{2.38}$$

Donde

$$\Sigma_{(0)}^{(i)} = \mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_0|\mathbf{y}_{1:K})}[(\mathbf{x}_0 - \mathbf{x}^b)(\mathbf{x}_0 - \mathbf{x}^b)^T], \tag{2.39}$$

$$\Sigma_{(k)}^{(i)} = \mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_{k-1}, \mathbf{x}_k|\mathbf{y}_{1:K})}[(\mathbf{x}_k - f_{k-1,k}(\mathbf{x}_{k-1}))(\mathbf{x}_k - f_{k-1,k}(\mathbf{x}_{k-1}))^T], \tag{2.40}$$

$$\Omega_{(k)}^{(i)} = \mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_k|\mathbf{y}_{1:K})}[(y_k - \mathbf{H}\mathbf{x}_k)(y_k - \mathbf{H}\mathbf{x}_k)^T] \tag{2.41}$$

Paso M

Descrita, para cada iteración i , la función intermedia $G_{\theta^{(i)}}(\theta)$ como en (2.38), se obtiene $\theta^{(i+1)}$ tomando el valor de θ que maximice esta función.

Para esto, derivamos respecto a los parámetros e igualamos a 0. Llamo \mathcal{E}_i a la $\mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_k|\mathbf{y}_{1:K})}$.

Para realizar las cuentas se debe tener en cuenta algunas propiedades para derivar respecto a los coeficientes de matrices:

$$(a) \quad \frac{\partial |A|}{\partial A} = |A| A^{-T}$$

$$(b) \quad \frac{\partial \ln|A|}{\partial A} = A^{-T}$$

$$(c) \quad \frac{\partial \text{tr}(A)}{\partial A} = I$$

$$(d) \frac{\partial \text{tr}(CA^{-1}B)}{\partial A} = -(A^{-1}BCA^{-1})^T = -A^{-T}C^T B^T A^{-T}$$

$$(e) \frac{\partial BAA^T}{\partial A} = BA + B^T A$$

Se deriva a continuación $G_{\theta^{(i)}}(\theta)$ con respecto a \mathbf{B} , \mathbf{Q} , \mathbf{R} y \mathbf{x}^b , en ese orden.

Derivar la igualdad (2.38) con respecto a \mathbf{B} es equivalente a derivar los primeros dos términos, ya que el resto es constante respecto a \mathbf{B} .

$$\begin{aligned} \frac{\partial G_{\theta^{(i)}}(\theta)}{\partial \mathbf{B}} &= \frac{\partial}{\partial \mathbf{B}} \left(-\frac{1}{2} \ln |\mathbf{B}| - \frac{1}{2} \text{tr} \{ \mathbf{B}^{-1} \Sigma_{(0)}^{(i)} \} \right) \\ &= -\frac{1}{2} \mathbf{B}^{-T} - \frac{1}{2} \mathbf{I} \frac{\partial}{\partial \mathbf{B}} (\mathbf{B}^{-1} \Sigma_{(0)}^{(i)}) \quad (\text{por (b) y (c)}) \\ &= -\frac{1}{2} \mathbf{B}^{-T} - \frac{1}{2} (-1) (\mathbf{B}^{-1} \Sigma_{(0)}^{(i)} \mathbf{B}^{-1})^T \quad (\text{por (d)}) \\ &= -\frac{1}{2} \mathbf{B}^{-T} + \frac{1}{2} \mathbf{B}^{-T} \Sigma_{(0)}^{(i)T} \mathbf{B}^{-T} \end{aligned} \tag{2.42}$$

Notar que en el segundo paso se agrega la matriz de manera auxiliar, y que $\Sigma_{(0)}^{(i)}$ es constante respecto de \mathbf{B} , así, es posible aplicar (d) para terminar de derivar.

Se continúa con la derivación respecto a \mathbf{Q} , la cual se resuelve de manera similar. De la igualdad (2.38) se derivan el tercer y cuarto término, y se utiliza que $\Sigma_{(j)}^{(i)}$ es constante respecto de \mathbf{Q} para todo j .

$$\begin{aligned} \frac{\partial G_{\theta^{(i)}}(\theta)}{\partial \mathbf{Q}} &= \frac{\partial}{\partial \mathbf{Q}} \left(-\frac{K}{2} \ln |\mathbf{Q}| - \frac{1}{2} \text{tr} \{ \mathbf{Q}^{-1} \sum_{j=1}^k \Sigma_{(j)}^{(i)} \} \right) \\ &= -\frac{K}{2} \mathbf{Q}^{-T} - \frac{1}{2} \mathbf{I} \frac{\partial}{\partial \mathbf{Q}} (\mathbf{Q}^{-1} \sum_{j=1}^k \Sigma_{(j)}^{(i)}) \quad (\text{por (b) y (c)}) \\ &= -\frac{K}{2} \mathbf{Q}^{-T} - \frac{1}{2} (-1) (\mathbf{Q}^{-1} [\sum_{j=1}^k \Sigma_{(j)}^{(i)}] \mathbf{Q}^{-1})^T \quad (\text{por (d)}) \\ &= -\frac{K}{2} \mathbf{Q}^{-T} + \frac{1}{2} \mathbf{Q}^{-T} [\sum_{j=1}^k \Sigma_{(j)}^{(i)}]^T \mathbf{Q}^{-T} \end{aligned} \tag{2.43}$$

Se deriva ahora respecto a \mathbf{R} . De la igualdad (2.38) se derivan los últimos dos términos, y se utiliza que $\Omega_{(j)}^{(i)}$ es constante respecto de \mathbf{R} para todo j .

$$\begin{aligned}
\frac{\partial G_{\theta^{(i)}}(\theta)}{\partial \mathbf{R}} &= \frac{\partial}{\partial \mathbf{R}} \left(-\frac{K}{2} \ln |\mathbf{R}| - \frac{1}{2} \text{tr} \left\{ \mathbf{R}^{-1} \sum_{j=1}^k \Omega_{(j)}^{(i)} \right\} \right) \\
&= -\frac{K}{2} \mathbf{R}^{-T} - \frac{1}{2} \mathbf{I} \frac{\partial G_{\theta^{(i)}}(\theta)}{\partial \mathbf{R}} \left(\mathbf{R}^{-1} \sum_{j=1}^k \Omega_{(j)}^{(i)} \right) \quad (\text{por (b) y (c)}) \\
&= \frac{K}{2} \mathbf{R}^{-T} + \frac{1}{2} \mathbf{R}^{-T} \left[\sum_{j=1}^k \Omega_{(j)}^{(i)} \right]^T \mathbf{R}^{-T} \quad (\text{por (d)}) \\
&= -\frac{K}{2} \mathbf{R}^{-T} + \frac{1}{2} \mathbf{R}^{-T} \left[\sum_{j=1}^k \Omega_j^{(i)} \right]^T \mathbf{R}^{-T}
\end{aligned} \tag{2.44}$$

Por último, se deriva respecto a \mathbf{x}^b . En este caso es equivalente a derivar sólo el segundo término de la igualdad (2.38).

$$\begin{aligned}
\frac{\partial G_{\theta^{(i)}}(\theta)}{\partial \mathbf{x}^b} &= \frac{\partial}{\partial \mathbf{x}^b} \left(-\frac{1}{2} \text{tr} \{ \mathbf{B}^{-1} \Sigma_{(0)}^{(i)} \} \right) \\
&= -\frac{1}{2} \mathbf{I} \frac{\partial}{\partial \mathbf{x}^b} \left(\mathbf{B}^{-1} \Sigma_{(0)}^{(i)} \right) \quad (\text{por (c)}) \\
&= -\frac{1}{2} \mathbf{B}^{-1} \frac{\partial}{\partial \mathbf{x}^b} \left(\mathcal{E}_{p_{\theta^{(i)}}(\mathbf{x}_0 | \mathbf{y}_{1:K})} [(\mathbf{x}_0 - \mathbf{x}^b)(\mathbf{x}_0 - \mathbf{x}^b)^T] \right) \\
&= -\frac{1}{2} \int \frac{\partial}{\partial \mathbf{x}^b} \left(\mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b)(\mathbf{x}_0 - \mathbf{x}^b)^T \right) p_{\theta^{(i)}}(\mathbf{x}_0 | \mathbf{y}_{1:K}) \partial \mathbf{x}_0 \\
&= -\frac{1}{2} \int \left(-\mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b) - (\mathbf{B}^{-1})^T (\mathbf{x}_0 - \mathbf{x}^b) \right) p_{\theta^{(i)}}(\mathbf{x}_0 | \mathbf{y}_{1:K}) \partial \mathbf{x}_0 \quad (\text{por (e)}) \\
&= \frac{1}{2} (\mathbf{B}^{-1} + \mathbf{B}^{-T}) \int (\mathbf{x}_0 - \mathbf{x}^b) p_{\theta^{(i)}}(\mathbf{x}_0 | \mathbf{y}_{1:K}) \partial \mathbf{x}_0 \\
&= \frac{1}{2} (\mathbf{B}^{-1} + \mathbf{B}^{-T}) \mathcal{E}_i(\mathbf{x}_0 - \mathbf{x}^b)
\end{aligned} \tag{2.45}$$

A continuación se igualan a cero todos los resultados generados para obtener la estimación para $\theta(i+1)$:

De (2.42) se encuentra la estimación de B en la iteración $i+1$:

$$\begin{aligned}
-\frac{1}{2} \mathbf{B}^{-T} + \frac{1}{2} \mathbf{B}^{-T} \Sigma_{(0)}^{(i)T} \mathbf{B}^{-T} &= 0 \\
\frac{1}{2} \mathbf{B}^{-1} \Sigma_{(0)}^{(i)} \mathbf{B}^{-1} &= \frac{1}{2} \mathbf{B}^{-1} \\
\mathbf{B}^{-1} \Sigma_{(0)}^{(i)} &= \mathbf{I} \\
\Sigma_{(0)}^{(i)} &= \mathbf{B}
\end{aligned} \tag{2.46}$$

De (2.43) se encuentra la estimación de \mathbf{Q} en la iteración $i + 1$:

$$\begin{aligned}
-\frac{k}{2}\mathbf{Q}^{-T} + \frac{1}{2}\mathbf{Q}^{-T}[\sum_{j=1}^k \Sigma_{(j)}^{(i)}]^T \mathbf{Q}^{-T} &= 0 \\
\frac{1}{2}\mathbf{Q}^{-1}[\sum_{j=1}^k \Sigma_{(j)}^{(i)}] \mathbf{Q}^{-1} &= \frac{k}{2}\mathbf{Q}^{-1} \\
[\sum_{j=1}^k \Sigma_{(j)}^{(i)}] &= k\mathbf{Q} \\
\frac{1}{k}[\sum_{j=1}^k \Sigma_{(j)}^{(i)}] &= \mathbf{Q}
\end{aligned} \tag{2.47}$$

De (2.44) se encuentra la estimación de \mathbf{R} en la iteración $i + 1$, de manera análoga a cómo se encontró la estimación de \mathbf{Q} en (2.47). Se obtiene:

$$\frac{1}{k} \sum_{j=1}^k \Omega_j^{(i)} = \mathbf{R} \tag{2.48}$$

Por último, de (2.45) se deriva la estimación de \mathbf{x}^b en la iteración $i + 1$:

$$\begin{aligned}
\mathbf{x}^b &= \mathcal{E}_i(\mathbf{x}_0) \\
\mathbf{x}^b &= \mathbf{x}_0^s
\end{aligned} \tag{2.49}$$

Así, una vez computadas las esperanzas $\Sigma_{(0)}^{(i)}$, $\Sigma_{(j)}^{(i)}$ y $\Omega_j^{(i)}$, y realizado un smoother, se estima $\theta(i + 1)$ de la siguiente manera:

$$\boxed{
\begin{aligned}
\hat{\mathbf{x}}^{b(i+1)} &= \mathcal{E}_i(\mathbf{x}_0) \\
\hat{\mathbf{B}}^{(i+1)} &= \Sigma_0^{(i)} \\
\hat{\mathbf{Q}}^{(i+1)} &= \frac{1}{k} \sum_{j=1}^k \Sigma_j^{(i)} \\
\hat{\mathbf{R}}^{(i+1)} &= \frac{1}{k} \sum_{j=1}^k \Omega_j^{(i)}
\end{aligned} \tag{2.50}$$

De esta manera hemos obtenido en forma analítica el paso de maximización a través de la hipótesis Gaussiana.

2.4.3. Ecuaciones y pseudocódigo para EM-EnKS

El problema de aproximar la pdf de $p_\theta(\mathbf{x}_k | \mathbf{y}_{1:K})$ es un problema de suavizado, para el cual se puede utilizar Suavizado de Kalman Extendido (Extended

Kalman Smoother, EKS) para estimar $\mathbf{x}_k^s, \mathbf{P}_k^s$ y computar $\mathcal{N}(\mathbf{x}_k^s, \mathbf{P}_k^s)$ como solución.

Se utiliza las Ecuaciones (2.24) para computar Σ_0 y Ω_k en (2.39) y (2.41). Para computar (2.40) se requiere conocer la covarianza cruzada, la cual no se calcula directamente del algoritmo EKS descrito, sino que se recurre a cuentas de Shumway y Stoffer (1982, [32]):

$$\begin{aligned} \mathbf{P}_{k,k-1}^s &\equiv \mathcal{E}_{p(\mathbf{x}_k, \mathbf{x}_{k-1} | y_{0:k})} \left[(\mathbf{x}_k - \mathbf{x}_k^s) (\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^s)^\top \right], \\ \mathbf{P}_{K,K-1}^s &= \left(\mathbf{I} - \mathbf{K}_K^f \mathbf{H} \right) \mathbf{M}_{K-1,K} \mathbf{P}_{k-1}^s. \end{aligned} \quad (2.51)$$

Luego, en el paso de maximización, se utilizan los resultados del EKS y (2.51) para computar (2.39)-(2.41) en cada iteración (i), computándose las covarianzas de la siguiente manera:

$$\begin{aligned} \Sigma_0^{(i)} &= \mathbf{P}_0^s + \left(\mathcal{E}_{p_\theta^{(i)}(\mathbf{x}_0 | y_{1:k})} [\mathbf{x}_0] - \mathbf{x}^b \right) \left(\mathcal{E}_{p_\theta^{(i)}(\mathbf{x}_0 | y_{1-\kappa})} [\mathbf{x}_0] - \mathbf{x}^b \right)^\top \\ &= \mathbf{P}_0^s \end{aligned} \quad (2.52)$$

$$\begin{aligned} \Sigma_k^{(i)} &= \mathbf{P}_k^s + \left\{ \mathbf{x}_k^s - f_{k-1,k}(\mathbf{x}_{k+1}^s) \right\} \left\{ \mathbf{x}_k^s - f_{k-1,k}(\mathbf{x}_{k+1}^s) \right\}^\top \\ &\quad + \mathbf{M}_{k-1,k} \mathbf{P}_{k-1}^s \mathbf{M}_{k-1,k}^\top - \mathbf{P}_{k-1,k}^s \mathbf{M}_{k-1,k}^\top - \mathbf{M}_{k-1,k} \mathbf{P}_{k-1,k}^s \end{aligned} \quad (2.53)$$

$$\Omega_k^{(i)} = (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k^s) (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k^s)^\top + \mathbf{H}\mathbf{P}_k^s \mathbf{H}^\top \quad (2.54)$$

Finalmente, para la media del estado de inicialización, se usa $\hat{\mathbf{x}}^{b,(i)} = \mathbf{x}_0^s$.

Ecuaciones EM para EnKS

En el caso que se utilice ensambles para el proceso de asimilación, las medias y matrices de covarianzas quedarán determinadas por las partículas de la muestra. A partir de las definiciones de (2.50), y de sus respectivas dependencias con (2.39), (2.40) y (2.41), se calculan las esperanzas, quedando determinadas las ecuaciones de la siguiente manera:

$$\begin{aligned} \hat{\mathbf{x}}^{b,(i)} &= \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{x}_{0,j}^s \\ \Sigma_0^{(i)} &= \frac{1}{N_e} \sum_{j=1}^{N_e} \left(\mathbf{x}_{0,j}^s - \hat{\mathbf{x}}^{b,(i)} \right) \left(\mathbf{x}_{0,j}^s - \hat{\mathbf{x}}^{b,(i)} \right)^\top \\ \Sigma_k^{(i)} &= \frac{1}{N_e} \sum_{j=1}^{N_e} \left\{ \mathbf{x}_{k,j}^s - f(\mathbf{x}_{k-1,j}^s) \right\} \left\{ \mathbf{x}_{k,j}^s - f(\mathbf{x}_{k-1,j}^s) \right\}^\top \\ \Omega_k^{(i)} &= \frac{1}{N_e} \sum_{j=1}^{N_e} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_{k,j}^s) (\mathbf{y}_k - \mathbf{H}\mathbf{x}_{k,j}^s)^\top \end{aligned} \quad (2.55)$$

N_x	Dimensión del estado del sistema.
N_e	Número de miembros de los ensambles.
K	Cantidad de observaciones, y de ciclos o tiempos de asimilación.
Δt	Distancia temporal entre dos observaciones consecutivas.
δt	Paso de tiempo del modelo, o de integración.
\mathbf{R}	Matriz de error de modelo de las observaciones.
\mathbf{Q}	Matriz de error aditivo del modelo.
\mathcal{H}	Operador de observaciones. \mathbf{H} si se lo toma lineal.
N_{iter}	Número de iteraciones

Tabla 2.1: Notación

Pseudocódigo de EM-EnKS

Se describe a continuación un pseudo-código para la implementación de EM-EnKS, ya que el objetivo final es utilizar este algoritmo para la estimación de parámetros. Como se describió anteriormente, el algoritmo EM permite estimar parámetros, tanto físicos como estadísticos, que maximicen la función de verosimilitud.

En Tandeo et al. (2015) se combinó este algoritmo con el filtro de Kalman por ensambles y se utilizó para estimar parámetros del operador observacional. En Dreano et al. (2016) se utilizó esta combinación con el objetivo de estimar la covarianza del error de modelo. En la presente tesina se utilizará para estimar el factor de inflación (aditivo y multiplicativo) de la covarianza de la predicción.

Para describir implementaciones, pseudocódigos y experimentos en esta tesina, se utilizará la notación que se presenta en la Tabla 2.1.

Luego, en el Algoritmo 1, se presenta un pseudocódigo para la implementación clásica del EM-EnKS, la cual es similar utilizada en Dreano et al. (2016).

Algorithm 1 EM

Input: $\mathbf{x}_{0,1:N_e}^a$, \mathbf{y}_0 , $\mathcal{M}(\cdot)$, θ_0 and $\text{filt}(\cdot)$

$\mathbf{x}_{0:K,1:N_e}^a$, $\mathbf{x}_{0:K,1:N_e}^f \leftarrow \text{filt}(\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, \theta_0)$ ▷ Filtro

$\mathbf{x}_{0:K,1:N_e}^s \leftarrow \text{suav}(\mathbf{x}_{0,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f)$ ▷ Suavizado

for $j = 1, N_{iter}$ **do**

$\theta_j \leftarrow \frac{1}{k} \sum_{j=1}^k \Sigma_j^{(i)}$ ▷ Estimación de θ

$\mathbf{x}_{0:K,1:N_e}^a$, $\mathbf{x}_{0:K,1:N_e}^f \leftarrow \text{filt}(\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, \theta_j)$

$\mathbf{x}_{0:K,1:N_e}^s \leftarrow \text{suav}(\mathbf{x}_{0:K,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f)$ ▷ Suavizado

end for

Output: $\theta_{0:N_{iter}}$

Capítulo 3

Metodología y primeros experimentos

En la presente tesis se propone desarrollar técnicas que permitan la estimación de parámetros que controlan la covarianza mediante estimación por máxima-verosimilitud. Se trabajará con tres métodos para maximizar la log-verosimilitud: la evaluación por grilla, Newton-Raphson y el Expectation-Maximization (EM). Los primeros dos optimizan la log-verosimilitud aproximada, en cambio, el EM optimiza la log-verosimilitud completa.

El sistema de asimilación combina información resultante de un modelo predictivo con observaciones. Por ende, está compuesto por un modelo predictivo, que representa la evolución del sistema, y por un método de asimilación. En esta tesis se utilizará como método de asimilación el filtro de Kalman por ensambles con la implementación conocida por perturbación de observaciones [22], descrito en 2.2.6, y como modelos predictivos se utilizarán los sistemas dinámicos de los modelos Lorenz 63 y 96 ([33], [34]). En este capítulo se describe estos sistemas dinámicos que se utilizan como prueba de concepto para la evaluación de las técnicas. Se detalla los experimentos que se realizarán, y se conduce y examina un primer conjunto de experimentos que resultan en la divergencia del filtro debido a una subestimación de la covarianza de la predicción y por ende motivan la necesidad de incluir error de modelo y factor de inflación en la covarianza de la predicción para la realización de la asimilación por ensambles, aspectos que serán abordados a través de técnicas de estimación por máxima-verosimilitud en los capítulos siguientes.

3.1. Modelos de pronóstico de sistemas dinámicos

3.1.1. Modelos

La asimilación de datos es una técnica que permite estimar, a través de observaciones indirectas parciales y ruidosas, el estado de un sistema que evoluciona en el tiempo cuya área de aplicación puede ser físico, biológico, geológico, de finanzas o de cualquier otra índole.

El mayor desafío lo representan sistemas dinámicos que son caóticos cuyo estado es intrínsecamente difícil de pronosticar. Las tres características principales de los sistemas caóticos son: la no-linealidad, la sensibilidad ante cambios pequeños de condiciones iniciales y que no se puede predecir el comportamiento del sistema con suficiente antelación.

De particular dificultad para la predicción es la caoticidad encontrada en los sistemas geofísicos, tanto atmosféricos como oceánicos debido a su alta dimensionalidad e imprevisibilidad [34].

En este sentido, para evaluar las técnicas que se desarrollarán se utilizarán modelos numéricos de predicción representando a sistemas dinámicos caóticos simples. En particular se utilizará el sistema caótico no-lineal Lorenz 1963 (L63) [33], de tres variables, y como modelo intermedio, de mayor dimensionalidad, el modelo de Lorenz 1996 (L96) de 40 variables [34]. Estos modelos no lineales y caóticos de pocas variables permiten evaluar las técnicas desarrolladas exhaustivamente con costos computacionales relativamente menores. En general estos modelos son usados como prototipos particularmente para aplicaciones en geociencias. El modelo de L96 ha sido creado para simular la variabilidad y predictibilidad de las variables atmosféricas [34].

Sistema de Lorenz 63

El sistema dinámico de Lorenz es un sistema de 3 ecuaciones que representa en forma simplificada los roles de convección que se producen en las ecuaciones dinámicas de la atmósfera terrestre. Es un sistema dinámico, determinista, tridimensional y no lineal. Este sistema dinámico, y el análisis de su atractor, fue introducido por Edward Lorenz en 1963. La forma del atractor de Lorenz se asemeja a las alas de una mariposa.

Para las variables (x, y, z) y constantes reales positivas σ, ρ y β , donde σ es el número de Prandtl y ρ es el número de Rayleigh, el sistema está determinado por las siguientes ecuaciones:

$$\begin{aligned}\frac{\partial x}{\partial t} &= \sigma(y - x) \\ \frac{\partial y}{\partial t} &= x(\rho - z) - y \\ \frac{\partial z}{\partial t} &= xy - \beta z\end{aligned}\tag{3.1}$$

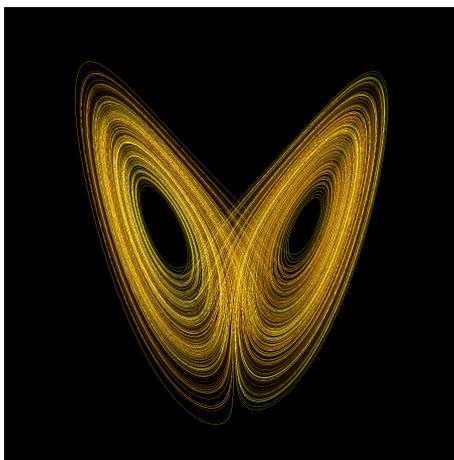


Figura 3.1: [Imagen ilustrativa tomada de Wikipedia] El atractor de Lorenz 63, con valores $\rho = 28$, $\sigma = 10$, $\beta = 8/3$.

Usualmente, se toma $\sigma = 10$, $\beta = 8/3$ y ρ variado. Para generar un sistema caótico se suele tomar $\rho = 28$. En la Figura 3.1 se observa la solución del sistema con estos valores siendo visible la estructura del atractor.

A los fines de ilustrar aspectos de la naturaleza caótica de este sistema, en la Figura 3.2 se muestra una trayectoria de referencia (Xt) del modelo Lorenz 63 durante 1000 tiempos. A su vez se muestran una trayectoria que resulta de una modificación (Xp) a través de pequeñas perturbaciones en las condiciones iniciales. Se realizan perturbaciones en los tiempos 0, 200, 400, 600 y 800. Las perturbaciones se realizan adicionando un vector aleatorio $w \sim \mathcal{N}(0; 0.01\mathbf{I})$, lo cual representa alteraciones del orden del 1% del rango de las variables. Al integrar el modelo a partir de estas perturbaciones se llega a alcanzar diferencias del 37% del rango de la variable en la variable X , y del 45% en la variable Y .

Se observa la sensibilidad del modelo a pequeñas perturbaciones. Dada esta sensibilidad, los modelos que se utilizan para representar y pronosticar los modelos de sistemas geofísicos complejos necesitan ser corregidos cada cierto (corto) periodo de tiempo incorporando información de las observaciones para no perder la representatividad de la realidad. La manera de corregirlos es utilizando -asimilando- información de observaciones del estado real.

Sistema de Lorenz 96

El modelo Lorenz 96 fue formulado, también por Edward Lorenz, en 1996. Es un modelo atmosférico unidimensional simple de disturbios que se propagan en un círculo de latitud, de naturaleza caótica, por lo que fue usado por Lorenz para demostrar problemas de predictibilidad atmosférica. Es un sistema dinámico no lineal, dado por un conjunto de N ecuaciones, definido de la siguiente manera

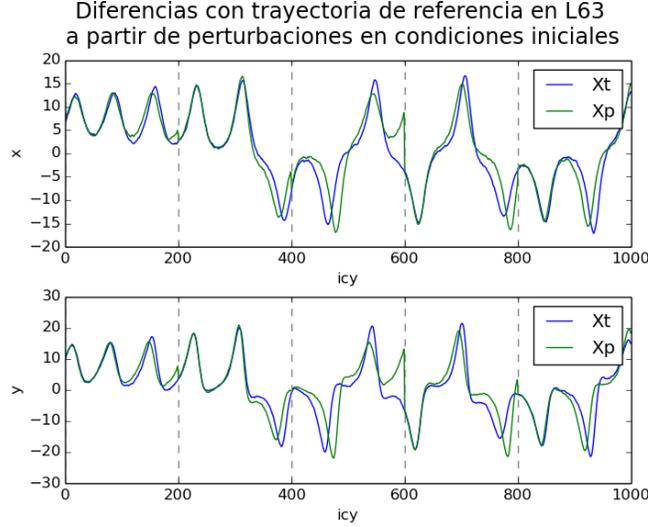


Figura 3.2: Modificación de una trayectoria de referencia del Modelo L63 (variables X e Y) a partir de pequeñas perturbaciones en diferentes tiempos.

por N variables x_i $i = 1, \dots, N$,

$$\frac{\partial x_i}{\partial t} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F. \quad (3.2)$$

Las variables se encuentran en un dominio periódico, es decir, se asume $x_{-1} = x_{N-1}$, $x_0 = x_N$ y $x_{N+1} = x_1$. F es una constante de forzado y representa un forzado externo (la energía proveniente del sol). Usualmente, se toma $F = 8$ para generar un sistema caótico, para valores menores del forzado el sistema será disipativo [35]. En la Figura 3.3 se observa una simulación de las primeras tres variables.

3.1.2. Configuración de parámetros para los experimentos

A lo largo de esta tesis se realizarán diferentes experimentos utilizando los modelos L63 y L96. Para los parámetros involucrados en estos modelos y en el sistema de asimilación de datos se utilizará la notación descrita en la Tabla 2.1.

Algunos de estos parámetros permanecen fijos a lo largo de los experimentos. En el caso de las simulaciones con el modelo Lorenz 63, se asumió que las observaciones fueron realizadas en forma secuencial con un periodo de 0.01 ($\Delta t = 0.01$), y que el paso de integración del modelo es $\delta t = 0.001$ es decir que entre una observación y otra se realiza 10 integraciones. Para la integración se utiliza el método de Runge Kutta de 4to. orden.

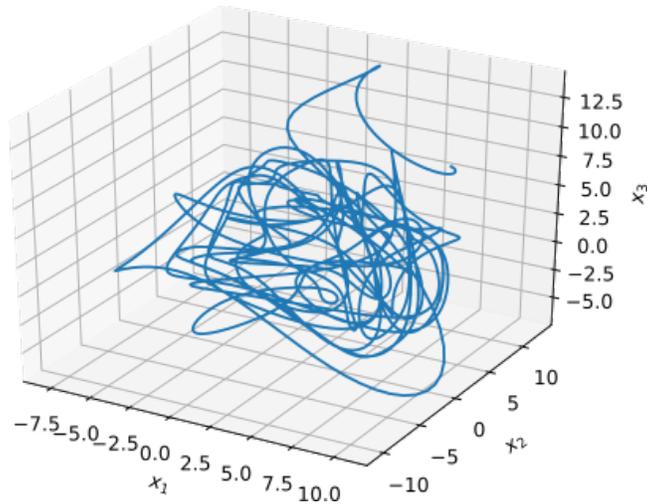


Figura 3.3: [Imagen ilustrativa tomada de Wikipedia] Evolución de las primeras tres variables de una simulación del Lorenz 96

En el caso de los experimentos que involucran el modelo Lorenz 96, se asumió que las observaciones fueron realizadas con una frecuencia de 0.05 ($\Delta t=0.05$), y que el modelo realiza 25 integraciones entre una observación y otra ($\delta t=\Delta t/25$). Además, se utilizará este modelo con 40 variables ($N_x=40$) siendo esta la configuración más usual. En ambos modelos se asumió que se están observando todas las variables del modelo.

3.2. Marco experimental

3.2.1. Experimentos gemelos - OSSE

Para estudiar los sistemas de asimilación y sus componentes se utilizan experimentos de simulación de sistemas de observación (observing system simulation experiments, OSSE) o también llamados experimentos gemelos. Estos son aquellos en los cuales se obtiene a través del sistema de ecuaciones el estado “verdadero” o real del sistema, y a partir de este, se generan observaciones sintéticas incluyendo el ruido de medición y de representación y simulando las características del tipo de instrumento de medición (variables observadas, frecuencia de observación, error observacional, etc.).

Utilizando estas observaciones sintéticas, se realiza la estimación del estado con el sistema de asimilación, como si no se conociera la verdad. Este tipo de experimentos permite evaluar el error de la estimación obtenida con el sistema de asimilación, ya que la verdad es conocida, y por ende se utiliza en experimentos de evaluación de técnicas, ya que en el marco operativo, el estado real no es

conocido, por lo que no es posible determinar el error de la estimación o si existen problemas, i.e. sesgos, en la estimación.

A continuación se detalla el procedimiento seguido en este tipo de experimentos. A lo largo de la tesis será utilizado este procedimiento, realizando variaciones en algunos de los pasos.

1. Se genera el estado real del sistema:

Se asume que se conoce exactamente la evolución del sistema. La evolución exacta se obtiene integrando un modelo numérico. Para las condiciones iniciales, se requiere comenzar con algún estado que ya se encuentre en el atractor del sistema. Para esto, se parte de un estado aleatorio, y a partir de este se integra el modelo por un largo tiempo (spin-up), por ejemplo, 5000 tiempos, considerando que después de esta integración el estado se encuentra en el atractor del sistema, se utiliza el estado resultante final (a los 5000 tiempos) como condición inicial del estado real.

Luego, a partir de esta condición inicial, se integra el modelo, y en cada tiempo, antes de volver a integrar, se agrega un error aleatorio que representa el error del modelo.

Sea:

- \mathcal{M}_t modelo que simula el avance real del estado
- t_s tiempos de spin-up
- \mathbf{x}_1 una realización de $\mathcal{N}(0, \mathbf{I}_{N_x \times N_x})$

Para $i = 1, \dots, t_s$ se genera:

- $\mathbf{x}_{i+1} = \mathcal{M}(\mathbf{x}_i)$
- (\mathbf{x}_{t_s} se asume que pertenece al atractor del modelo)

Se define la condición inicial:

- $\mathbf{x}_1^t = \mathbf{x}_{t_s}$, simulación del estado real (true, representado con supraíndice t) en tiempo 0

Luego, se integra y simula el estado real

- $\mathbf{x}_{k+1}^t = \mathcal{M}(\mathbf{x}_k^t) + \eta_k$, $k = 1, \dots, K$, donde $\eta_k \sim \mathcal{N}(0, \mathbf{Q}^t)$,
- simulación del estado real en el tiempo $i + 1$ donde \mathbf{Q}^t es la covarianza del error de modelo.

2. Se generan observaciones sintéticas:

A partir la obtención del estado real, se generan las observaciones sintéticas, esto es, se toma el valor de algunas variables a tiempos determinados de la evolución exacta y estas son proyectadas al espacio de observaciones. En sistemas espaciales se debe considerar que los puntos de observación pueden no coincidir con los puntos de grilla del modelo (ej. mediciones satelitales). Luego, se adiciona un error aleatorio de una dada distribución. Este error representa tanto el error instrumental como el error de representación.

Para $k = 1, \dots, K$, se genera:

$$- \mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \epsilon_k, \quad \text{donde } \eta_k \sim \mathcal{N}(0, \mathbf{R}).$$

3. Se genera un estado (o ensamble) inicial:

Existen al menos dos formas usuales de generar un ensamble de estados iniciales requerido por el sistema de asimilación. Una, la que se implementó en esta tesis, y es análoga a la que se generó el estado inicial, es realizando una integración larga, de por ejemplo 5000 tiempos eliminando los primeros tiempos del spinup. A partir de estos se computa la covarianza climatológica como la covarianza de estos estados, y además, se elige alguno de los estados para usar como media. Se toma una muestra de n estados a partir de la densidad Gaussiana dada por esta media y covarianza, los cuales conformarán el ensamble inicial.

Se detalla a continuación el procedimiento. En este se elige el último estado del spin-up como media, pero podría ser cualquiera:

Sea:

- \mathbf{x}_1 una realización de $\mathcal{N}(0, \mathbf{I}_{N_{xx}, N_x})$, independiente de la anterior.
- $\mathbf{x}_{i+1} = \mathcal{M}(\mathbf{x}_i)$, $i = 1, \dots, t_s$.
- $\hat{\mathbf{x}}_1 = \mathbf{x}_{t_s}$, será la media del ensamble inicial

Luego, sea:

$$- \mathbf{x}_m = \frac{1}{t_s} \sum_1^{t_s} \mathbf{x}_i$$

$$- \hat{\mathbf{P}}_1 = \frac{1}{t_s} \sum_j^{t_s} (\mathbf{x}_i - \mathbf{x}_m)(\mathbf{x}_i - \mathbf{x}_m)^T, \quad \text{covarianza del ensamble inicial.}$$

Se define:

- $\mathbf{x}_{1,j}^a$, realización j de $\mathcal{N}(\hat{\mathbf{x}}_1, \hat{\mathbf{P}}_1)$, con $j = 1, \dots, N_e$.

Por último:

- $\{\mathbf{x}_{1,j}^a\}_{j=1}^{N_e}$ será el ensamble inicial.

Alternativamente, se realiza el spin-up a partir de un estado aleatorio, y luego se continúa integrando el sistema por un plazo de tiempo relativamente largo (10.000 tiempos). Tomándose los N_e estados requeridos para el ensamble en forma aleatorio del conjunto de 10.000 estados (se asume que los miembros del ensamble no están correlacionados temporalmente).

4. Se aplica el sistema de asimilación:

Se aplica el sistema de asimilación a estas observaciones sintéticas. Es

decir, en cada tiempo se pasará el filtro combinando un modelo con las observaciones dadas para ese tiempo.

Para cada tiempo k , $k = 1, \dots, K$:

Paso de pronóstico (ec.(2.22)):

$$- \mathbf{x}_{k,j}^f = \mathcal{M}^f(\mathbf{x}_{k-1,j}^a) + \eta_{k,j}, j = 1, \dots, N_e, \text{ con } \eta_{k,j} \sim \mathcal{N}(0, \mathbf{Q})$$

$$- \mathbf{y}_{k,j}^f = \mathcal{H}\left(x_{k,j}^f\right) + \epsilon_{k,j}, j = 1, \dots, N_e, \text{ con } \epsilon_{k,j} \sim \mathcal{N}(0, \mathbf{R})$$

Paso de análisis (ec.(2.23)):

$$- \mathbf{x}_{k,j}^a = \mathbf{x}_{k,j}^f + \mathbf{K}_k^f \left(\mathbf{y}_k - \mathbf{y}_{k,j}^f \right), j = 1, \dots, N_e$$

Donde \mathcal{M}^f es el modelo con el que se realiza la asimilación, $\mathbf{y}_{k,j}^f$ es la proyección al espacio de las observaciones, en el tiempo k , del miembro j del ensemble y \mathbf{K}_k^f es la matriz de peso del filtro de Kalman (Ec. (2.20)).

En muchos casos, cuando se quiere asimilar observaciones en sistemas complejos y estimar el estado del sistema, solo es posible un modelo de predicción simplificado, o subrogado. De acuerdo al tipo de modelo de predicción, clasificamos a los experimentos en asimilación con modelo perfecto o modelo imperfecto.

Asimilación con modelo perfecto:

Es el caso en que el modelo numérico en el sistema de asimilación sea el mismo que el utilizado para generar las observaciones sintéticas ($\mathcal{M}^f = \mathcal{M}$). En este caso se está cumpliendo la hipótesis de que no existen errores sistemáticos en el modelo, por lo cual, estos experimentos gemelos permiten evaluar la técnica desarrollada en un ambiente ideal y controlado donde se satisfacen (parcialmente) las hipótesis de la técnica (a excepción de la gaussianidad y linealidad).

Asimilación con modelo imperfecto:

En estos casos el modelo de predicción en el sistema de asimilación es una simplificación del sistema real. Para simular este tipo de entorno en experimentos de prueba de concepto se puede utilizar un modelo de predicción con modificaciones al modelo real, por ejemplo, cambiando parámetros del modelo o cambiando la resolución del modelo. De esta forma se puede evaluar el potencial de la técnica bajo la presencia de error de modelo (aun cuando las observaciones sean sintéticas). Cabe notar que en este caso existe un sesgo sistemático en las predicciones del modelo, por lo que se está evaluando al sistema cuando una de las hipótesis del filtro no se cumplen, simulando situaciones reales en la cual el modelo de predicción posee simplificaciones que conllevan a la presencia de sesgos y/o errores aleatorios del modelo.

3.2.2. Descripción de los tipos de experimentos realizados

Para realizar evaluaciones del sistema de asimilación en los entornos mencionados en la subsección anterior, en esta tesis se realizaron experimentos que

asimilan con modelo perfecto o imperfecto, y utilizando los sistemas de Lorenz 63 y de Lorenz 96, de esta manera definimos los experimentos por:

- **L63-per:** se generan las observaciones sintéticas, y se asimila, con el modelo Lorenz 63 con los parámetros usuales en simulaciones de sistemas caóticos: $\sigma = 10, \beta = 8/3, \rho = 28$ (utilizados tanto en el sistema real como en el modelo de predicción).
- **L63-imper:** se generan las observaciones sintéticas, y se asimila, con el modelo Lorenz 63. En la generación de observaciones se utilizan los parámetros usuales, $\sigma = 10, \beta = 8/3, \rho = 28$, y en la asimilación se cambia por $\sigma = 11,5, \beta = 2,87$ y $\rho = 32$ (sistema de asimilación con un modelo predictivo con error paramétrico).
- **L96-per:** se generan las observaciones sintéticas, y se asimila, con el modelo Lorenz 96 de 40 variables con el forzado usual: $F = 8$.
- **L96-imper:** se generan las observaciones sintéticas, y se asimila, con el modelo Lorenz 96 de 40 variables. En la generación de observaciones se utiliza el parámetro usual, $F = 8$, y en la asimilación se lo cambia por $F = 10$.

En secciones siguiente se hará referencia a experimentos del tipo perfecto o imperfecto, basándose en esta nomenclatura. Por otro lado, para representar el error de modelo y el error de muestreo en esta tesis se trabajará con experimentos que amplifiquen la incerteza del error de pronóstico en el sistema de asimilación. Se propone una combinación lineal de las matrices de covarianzas \mathbf{Q} como \mathbf{P}^f . Se utilizarán parámetros β y α para amplificar estos errores, respectivamente, como se describió en la Sec. 2.2.8.

El factor α se denomina comúnmente factor de inflación multiplicativo, y a β factor de inflación aditivo. Por ende, la matriz de covarianza de la predicción efectiva después de estas modificaciones viene dada por:

$$\mathbf{P}^f = \alpha \hat{\mathbf{P}}^f + \beta \mathbf{Q}$$

donde $\hat{\mathbf{P}}^f$ es la covarianza muestral determinada con los estados resultantes de la predicción por el modelo dinámico. Al introducir estos parámetros se modifica el sistema de asimilación a utilizar:

Con inflación multiplicativa:

Se introduce el factor α y se actualizan los miembros del ensamble del pronóstico, como se detalló en la ec. (2.25), antes de aplicarse el paso de análisis:

$$\mathbf{x}_{k,j}^f \leftarrow \bar{\mathbf{x}}_k^f + \sqrt{\alpha} \left(\mathbf{x}_{k,j}^f - \bar{\mathbf{x}}_k^f \right)$$

Actualizándose así la matriz de covarianza $\hat{\mathbf{P}}^f$ a $\alpha \hat{\mathbf{P}}^f$.

Con inflación aditiva:

Se introduce el factor β y se utiliza la matriz de error de modelo $\beta \mathbf{Q}$ en lugar

de \mathbf{Q} , en el caso de que se conozca o asuma una estimación de \mathbf{Q} . Si no se conoce, se utilizará un múltiplo de la identidad, quedando representado el error de modelo como $\mathbf{Q} = \beta \mathbf{I}$. En los casos de sistemas estocásticos la verdad es simulada usando un error de modelo parametrizado por un factor β^t .

A base de esto, la nomenclatura de los experimentos descrita anteriormente se le adiciona:

- **sininfla**: Sin inflación. No se amplifica la incertidumbre. Se utilizan los parámetros por defecto $\alpha = 1$ y $\beta = 0$.
- **infl-mult**: Con inflación multiplicativa. Se amplifica la matriz de covarianza del ensamble. Se utiliza α mayor a 1 el cual debe ser definido. Se prescribe $\beta = 0$.
- **infl-add**: Con inflación aditiva. Se amplifica la matriz de error de modelo. Se utiliza $\mathbf{Q} = \beta \mathbf{Q}_f$, pudiendo ser \mathbf{Q}_f la matriz identidad. Se debe definir este parámetro β , y se toma $\alpha = 1$.
- **infl-mult-y-add**: Con inflación multiplicativa y aditiva. Se amplifican ambas matrices de covarianza. Se deben definir o estimar: α y β .

El estado real es generado de manera determinística ($\mathbf{Q} = 0$) en los experimentos **sininfla** y **infl-mult**, y de manera estocástica ($\mathbf{Q} \neq 0$) en los otros casos. Esto se especifica en cada experimento.

3.3. Experimentos de referencia: motivación para el factor de inflación

En esta sección se realizan experimentos que motivan la introducción de parámetros de la covarianza en el sistema de asimilación, en particular se muestra como una sub representación de la incerteza de la predicción ocasiona en experimentos sencillos una divergencia del filtro, es decir el filtro termina estimando valores cercanos a la predicción y no incluye información observacional. Este proceso de divergencia del filtro es un proceso de retroalimentación, si en un paso se subestima (con respecto al valor óptimo cualquiera sea la definición de este) la incerteza de la predicción, la combinación lineal de la observación con la predicción dará un valor relativamente más cercano a la predicción, pero además la incerteza del análisis será subestimada. Esto conlleva que al realizar la predicción en el próximo tiempo observacional, la incerteza de predicción sea doblemente subestimada, debido a la disminución de la incerteza del paso anterior y de la propia predicción, por lo que la aplicación en forma secuencial de esto conlleva al colapso de la incerteza del estado, y, por lo tanto, la divergencia del filtro.

Se realizan primero, experimentos de modelo perfecto (como se describe en la Sección 3.2.2), en los que se observa que el filtro converge. Luego se realizan experimentos de modelo imperfecto en las mismas condiciones del filtro y se

muestra la divergencia del filtro. Por último, se incorpora un factor de inflación adecuado en este último marco y se observa como al amplificar la incerteza del modelo por medio de este factor, ahora, el filtro converge. Este impacto del error del modelo en el filtro se muestra evaluando tanto la media de los ensambles involucrados en los experimentos como la covarianza de los mismos. Estos experimentos van a ser usado como referencia para el resto de los experimentos en la tesis.

La inclusión de este factor de inflación multiplicativa es en principio arbitrario e incluido ad-hoc. En este sentido, un factor de inflación cercano a uno conllevaría nuevamente a la divergencia del filtro (como sucede en el caso estándar sin inflación) mientras si el factor es demasiado elevado, se está menospreciando al modelo de predicción y, por lo tanto, la incerteza de la estimación es elevada, por lo que es necesario encontrar estadísticamente una covarianza de la predicción que sea la óptima. Como se ha mencionado este es el objetivo de la tesis, desarrollar metodologías basadas en el principio de máximo verosimilitud, de un conjunto de observaciones distribuidas en el tiempo, que permitan estimar los factores de la covarianza de predicción y de la covarianza del error de modelo que maximizan la verosimilitud y, por lo tanto, optimicen la estimación. Es importante remarcar que nuestra propuesta para el objetivo asume que las observaciones en cada tiempo son limitadas por lo que consideramos las metodologías de máximo verosimilitud basadas en un batch de observaciones de extrema relevancia. El batch debe ser lo suficientemente extenso como para restringir a los parámetros estadísticos, i.e. a las matrices de covarianza, más allá de las variables de estado estimadas ciclo a ciclo.

Para los experimentos de esta subsección se asumió que se cuenta con 1000 observaciones y se generan ensambles de 100 miembros ($K=1000$ y $N_e=100$). Se asumió además error de observaciones $\mathbf{R} = 1.5\mathbf{I}_d$, y error de modelo $\mathbf{Q} = 0$.

3.3.1. Experimentos con modelo perfecto

Experimento 3.1: Series de tiempo y proceso de asimilación a partir del modelo L63 en marco de modelo perfecto sin factor de inflación.

Se realizó el experimento de modelo perfecto generando observaciones sintéticas a partir del modelo Lorenz 63, y luego se utiliza este mismo modelo como pronóstico y se asimilan esas observaciones. En el sistema de asimilación no se usa inflación. Es decir, un experimento del tipo L63-per-sininfla, como se describe en la Sección 3.2.2. Se muestra en la Figura 3.4 el estado real (X_t), las observaciones simuladas (Y_o), y las medias de los ensambles de pronóstico (X_f) y asimilación (X_a), obtenidos en este experimento. En cada panel se muestra la evolución temporal de cada una de las tres variables del modelo.

En este experimento se obtiene una muy buena performance del filtro, como se muestra el filtro es capaz de converger al valor real y luego eliminar el ruido observacional.

Experimento 3.2: Series de tiempo y proceso de asimilación a partir del

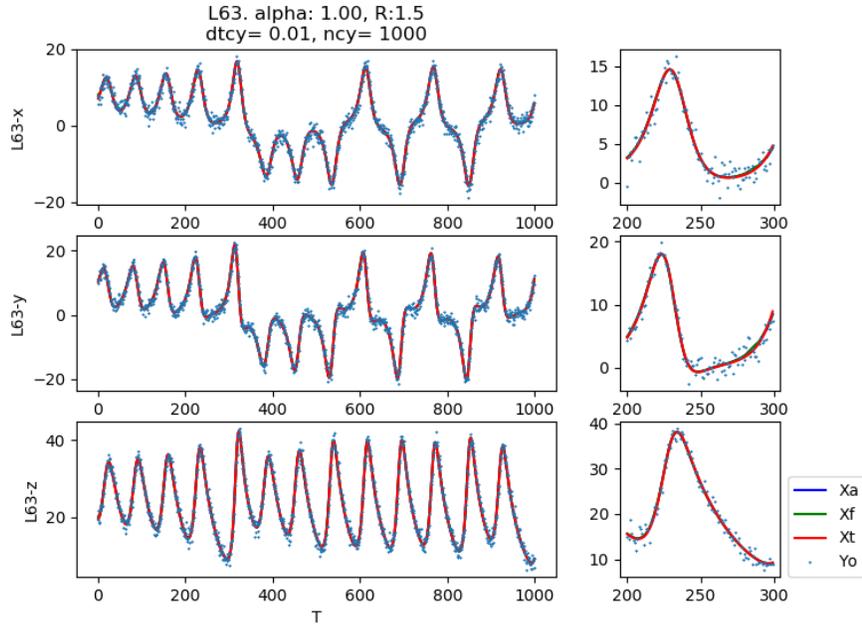


Figura 3.4: Series de tiempo y proceso de asimilación a partir del modelo L63 en marco de experimento perfecto sin factor de inflación. Experimento 3.1.

modelo L96 en marco de modelo perfecto sin factor de inflación.

De manera análoga se realizó un experimento de modelo perfecto con el modelo Lorenz 96. Específicamente, se realizó un experimento del tipo L96-per-sininfla.

Se muestra en la Figura 3.5 las observaciones y el estado real simulados, y las medias de los ensambles de pronóstico y la estimación dada por el sistema de asimilación, generados en este experimento, para las variables 0, 15 y 30.

Se obtiene una buena performance del filtro, el cual estima valores cercanos a la verdad.

3.3.2. Experimentos con modelo imperfecto

Modelo L63

Experimento 3.3: Series de tiempo y proceso de asimilación a partir del modelo L63 en marco de modelo imperfecto sin factor de inflación.

Se realizó el experimento de generar observaciones sintéticas a partir del

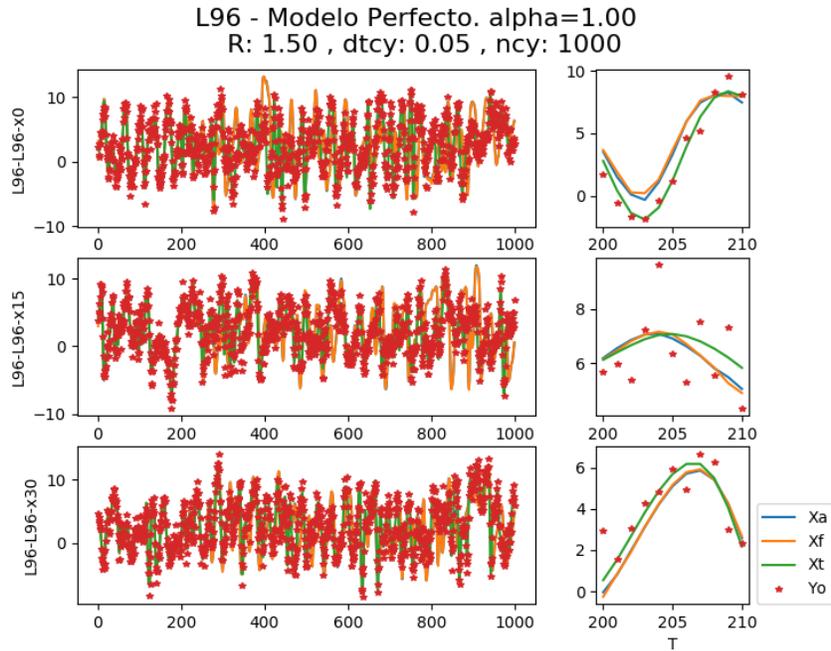


Figura 3.5: Series de tiempo y proceso de asimilación a partir del modelo L96 en marco de modelo perfecto sin factor de inflación. Experimento 3.2.

modelo Lorenz 63, configurado con los parámetros clásicos, y luego en el sistema de asimilación se utiliza un modelo Lorenz 63 imperfecto el cual es configurado con otros parámetros. Es decir, un experimento del tipo L63-imper, como se describe en la Sección 3.2.2. Se asimiló sin inflación.

Se muestra en la Figura 3.6 el estado real (X_t), las observaciones simuladas (Y_o), y las medias de los ensambles de pronóstico (X_f) y la estimación del análisis (X_a), obtenidos en este experimento.

En este experimento el filtro diverge, ya que se encuentran las observaciones cerca del estado real, por un lado, y por otro, la media del ensamble de análisis y la media del ensamble del pronóstico que son casi idénticas. Es decir, el filtro estima valores cercanos al pronóstico, y termina ignorando las observaciones, esto resulta de una subestimación de la incerteza de predicción por no considerar el error de modelo que conlleva la presencia del sesgo sistemático (debido al error paramétrico).

En un sistema de asimilación se tienen los ensambles de pronóstico y de análisis, y las medias de cada uno podría considerarse una estimación del estado real, esperando que la del análisis sea una mejor estimación. Si el sistema de asimilación funciona correctamente, es decir, si el filtro converge, se espera que

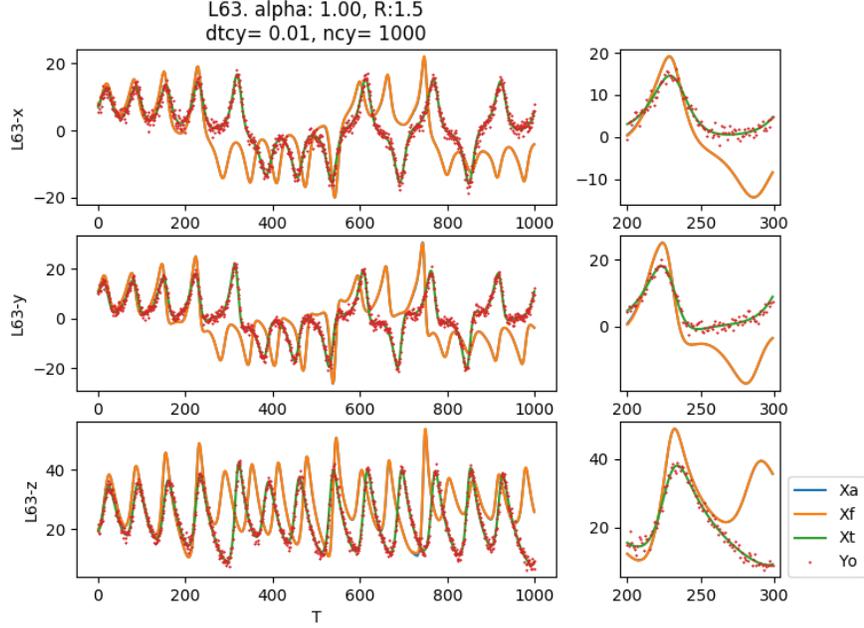


Figura 3.6: Series de tiempo y proceso de asimilación a partir del modelo L63 en marco de modelo imperfecto sin factor de inflación. Experimento 3.3.

la dispersión de los ensambles (que son la incerteza de las estimaciones) sean del orden del error de estimación de las medias de los mismos, por ende, se espera que las dispersiones, cuantificadas de alguna manera, sean de una escala comparable al RMSE entre la media de los ensambles y el estado real. Cabe aclarar que este último es factible de computar solo porque estamos en el marco de un experimento gemelo, donde se conoce el estado real.

Para medir la dispersión de los ensambles, se basa en el análisis multivariado, en el cual una medida de dispersión utilizada comúnmente es la traza de la matriz de covarianza. En este caso, se considerarán las trazas normalizadas de las matrices de covarianza de los ensambles del pronóstico y del análisis.

Por otro lado, para computar el RMSE entre la media del/los ensambles con el estado real, se realiza para cada tiempo de la siguiente manera:

Para cada tiempo k , con $k = 1, \dots, K$, se tiene el estado real generado \mathbf{x}_k^t y la media del ensamble de pronóstico $\bar{\mathbf{x}}_k^f$ (idem con $\bar{\mathbf{x}}_k^a$), vectores de dimensión N_x . Fijamos k y denominamos a estos vectores como v y \hat{v} respectivamente. El RMSE entre los vectores v y \hat{v} se calcula de la siguiente manera:

$$rmse(v, \hat{v}) = \sqrt{\frac{\sum_{l=1}^{N_x} (v(l) - \hat{v}(l))^2}{N_x}}$$

En la Figura 3.7 se muestra, para este mismo experimento, en el panel de la izquierda, para cada tiempo de asimilación la traza de las matrices de covarianza de los ensambles de pronóstico ($tr(P_f)$) y análisis ($tr(P_a)$); y en el panel de la derecha, se muestra para cada tiempo de asimilación, el error cuadrático medio (RMSE) entre la verdad generada, y la media de los ensambles de pronóstico ($rmse_{x_f}(t)$) y análisis ($rmse_{x_a}(t)$).

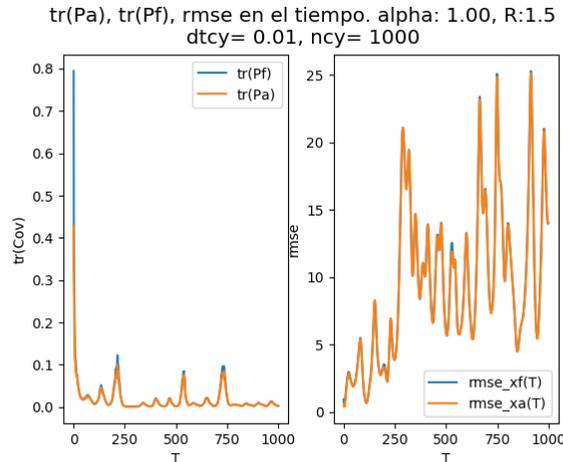


Figura 3.7: RMSE y trazas de matrices de covarianza en función del tiempo de asimilación a partir del modelo L63 sin factor de inflación. Experimento 3.3.

Notas:

- Las trazas de las matrices de covarianzas involucradas fueron reescaladas con la dimensión del estado (se dividió por la dimensión). Se obtuvo así el promedio de las varianzas muestrales de cada variable, y este valor se lo compara con el RMSE. Por simplicidad nos referiremos directamente como "Traza de matrices de covarianza".
- Por otro lado, cabe aclarar que al utilizar el RMSE en estos experimentos se está utilizando la verdad para estudiar el impacto del factor de inflación. Esto no es posible en un experimento real. Las técnicas que se describen en los siguientes capítulos solo usan las observaciones para determinar el factor de inflación.

Se puede ver como la traza de las matrices de covarianza dan en el orden de 0.02 y el RMSE en el orden de 5. Las magnitudes que toman estas cantidades son bien diferentes, entendiendo así que la traza de las matrices de covarianza no está representando la incerteza de la estimación (hay una importante subestimación). Se puede observar con esto que el ensamble está colapsando, ya que la traza tiende a cero, mientras que el error de estimación aumenta. Es decir, el ensamble colapsa en un punto lejano al estado real: el filtro divergió.

Viendo entonces que el ensemble está colapsando, se decide agregar un factor multiplicativo (Sec. 2.2.8) en la matriz de covarianza del filtro para ver el impacto en la asimilación. Se busca evitar que el ensemble colapse, y por ende, que el filtro diverja.

Experimento 3.4: Series de tiempo y proceso de asimilación a partir del modelo L63 en marco de modelo imperfecto con factor de inflación.

Se realizó un experimento del tipo L63-imper-alpha, con un valor de $\alpha = 1.78$ (Sec. 3.2.2). Es decir, se infla la covarianza del ensemble del pronóstico los miembros del ensemble del pronóstico con este factor, como se explicita en la ec. (2.25), antes de ser combinados con las observaciones para calcular el ensemble de análisis. Esto se realiza en cada tiempo de asimilación.

Con la incorporación de este factor se busca aumentar el error de la predicción y que este represente el error generado por el modelo, y lograr así que la media del ensemble de asimilación se aproxime mejor a las observaciones.

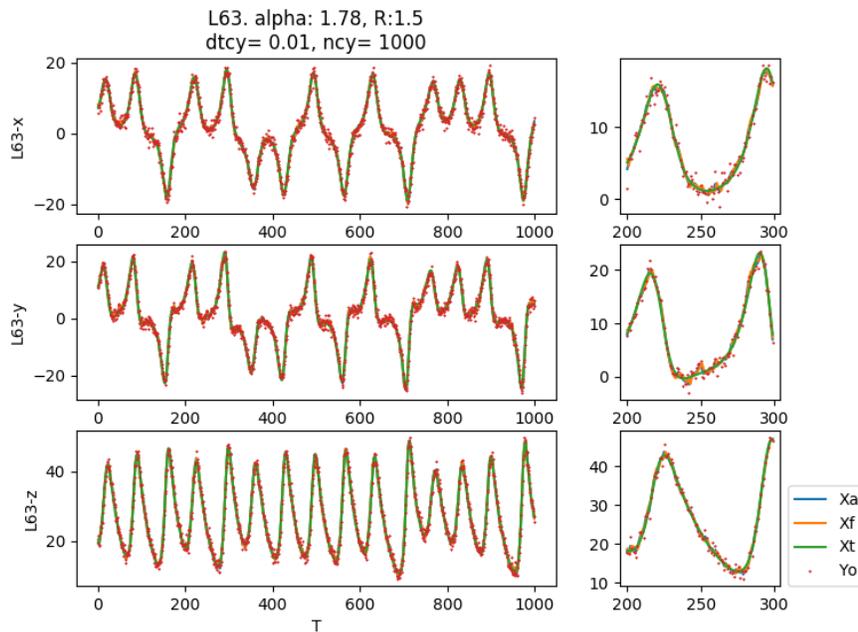


Figura 3.8: Series de tiempo y proceso de asimilación a partir del modelo L63 en marco de modelo imperfecto con factor de inflación. Experimento 3.4.

Se muestra en la Figura 3.8 las series de tiempo resultantes de este experimento. Se grafica nuevamente las series de tiempo para cada variable, el estado real simulado (X_t), las observaciones (Y_o), y las medias de los ensembles de

pronóstico (X_f) y las medias de los análisis (X_a).

Se observa como al incorporar el factor de inflación, se pasó de un escenario donde el filtro divergía, como en el Experimento 3.3 representado en la Figura 3.6, en donde la media del análisis y pronóstico se alejan de las observaciones; a un escenario donde el filtro converge, como en el Experimento 3.1 representado en la Figura 3.4, en donde la media de estos ensambles son buenos estimadores del estado real X_t .

Por ende, con la incorporación del factor de inflación logra representar más adecuadamente el error del modelo (error paramétrico en este caso), y corregir la divergencia del filtro, logrando así, con un modelo con error, una estimación del estado real cercana a las observaciones.

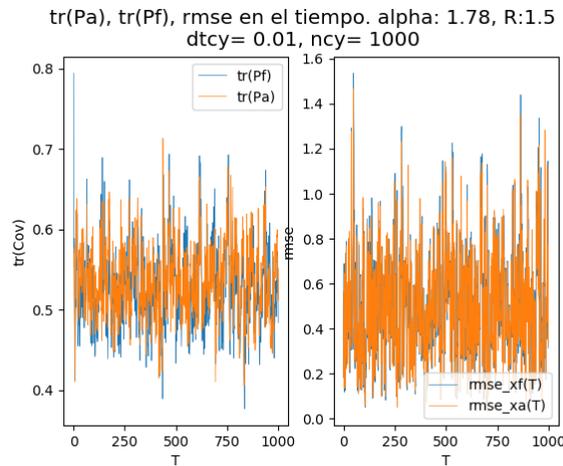


Figura 3.9: RMSE y trazas de matrices de covarianza en función del tiempo de asimilación a partir del modelo L63 con factor de inflación. Experimento 3.4.

Se muestra en la Figura 3.9 los RMSE y trazas de las matrices de covarianza de este experimento. Se observa que si bien parecen ser más variable los RMSE que las trazas de las matrices de covarianzas, ya que toma valores en el rango de 0.2 a 1.6 aproximadamente, y las trazas de las matrices entre 0.40 y 0.70, la media de ambos se asemejan a valores cercanos a 0.50 en ambos casos. Se muestra como el orden de estas variables ahora si son comparables, pudiendo ser así la traza de las matrices de covarianza un posible estimador del RMSE.

Esta Figura muestra otra evidencia de como el problema de la divergencia del filtro puede ser subsanada al incorporar un factor de inflación adecuado. Se muestra modo resumen, para observar valores precisos, en la Tabla 3.1 la media a lo largo de los ciclos de asimilación tanto de la traza de las matrices como del RMSE, en el marco de modelo imperfecto, con y sin inflación.

$\mathbf{R} = 1.5\mathbf{I}$		tr(Cov)	RMSE
Sin inflación	xf	0.0150	6.750
	xa	0.0142	6.682
Con inflación (1.46)	xf	0.525	0.697
	xa	0.535	0.578

Tabla 3.1: Promedio a lo largo de los ciclos de asimilación de las trazas de las matrices de covarianza y del RMSE a partir del modelo L63 en marco de modelo imperfecto, con y sin inflación. Experimentos 3.3 y 3.4.

Modelo L96

Se muestra a continuación experimentos y resultados análogos a los anteriores, pero utilizando el modelo Lorenz 96 en vez de Lorenz 63.

Experimento 3.5: Series de tiempo y proceso de asimilación a partir del modelo L96 en marco de modelo imperfecto sin factor de inflación.

Análogo al experimento 3.2, se realiza un experimento para el modelo Lorenz 96 del tipo L96-imper sin inflación. Se muestra en la Figura 3.10 los resultados del mismo para los primeros 400 tiempos de asimilación. Se muestra también en este caso, la divergencia del filtro. La media del ensamble de análisis casi coincide con la media del ensamble del pronóstico, ignorando la información de las observaciones. Se observa como el filtro no logra representar de manera adecuada el error del modelo, ya que le da mucho peso a este, y no logra estimar adecuadamente el estado real.

Del mismo modo se grafican las trazas de las matrices de covarianza y el RMSE. Se muestra en la Figura 3.11 estos resultados para cada tiempo. Se observa que la traza de las covarianzas toman valores en el orden de 0.08 mientras que el RMSE toma valores del orden de 5. Son valores del orden parecido a los mostrados para el modelo L63 en la Figura 3.7. Se puede deducir también que en este caso la traza de las matrices de covarianza de los ensambles no son buenos estimadores del RMSE cometidos.

Experimento 3.6: Series de tiempo y proceso de asimilación a partir del modelo L96 en marco de modelo imperfecto con factor de inflación.

Análogo al experimento 3.4, se realizó con el modelo Lorenz 96 un experimento del tipo imperfecto, pero con factor de inflación, es decir, experimento del tipo L96-imper-alpha (Sec. 3.2.2). Se utilizó como parámetro multiplicativo $\alpha = 1.67$.

La Figura 3.12 muestra el resultado de este experimento, para las variables 0, 15 y 30 en los primeros 400 tiempos de asimilación. Se observa como al incorporar el factor de inflación, se pasó del escenario de divergencia de filtro del Experimento 3.5 representado en la Figura 3.10, al de convergencia del filtro

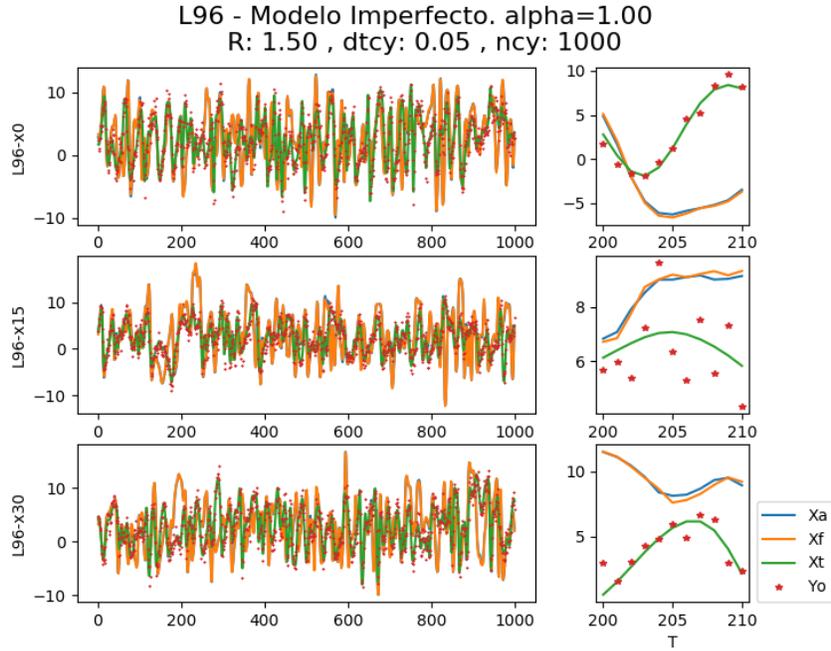


Figura 3.10: Series de tiempo y proceso de asimilación a partir del modelo L96 en marco de modelo imperfecto sin factor de inflación. Experimento 3.5.

similar al Experimento 3.2 representado en la Figura 3.5. El factor de inflación en este experimento representa de manera simplificada el error cometido por asimilar con un modelo simplificado.

En la Figura 3.13 se grafican para este Experimento las trazas de las matrices de covarianza y los RMSE para media de los ensambles. Se observa la similitud de los valores tomados por estas cantidades, a diferencia de la Figura 3.11, en donde tomaban valores de escalas distintas. Mostrando que la covarianza está representando (en orden de magnitud) la incerteza de la estimación. Se muestra en la Tabla 3.2, para poder comparar valores precisos, la media a lo largo de los ciclos de asimilación tanto de la traza de las matrices como del RMSE, con y sin inflación, para el modelo L96.

Es importante aclarar que es de gran influencia la magnitud de α que se utilice, ya que una subestimación conlleva a la divergencia del filtro que se mencionó mientras una sobreestimación asigna una covarianza demasiado grande a las predicciones resultando en un filtro con mayor incerteza. Se trabajará, por tanto, en esta tesis en métodos para encontrar el valor óptimo de este factor de inflación, como así también del factor β en el caso de trabajar con inflación aditiva, o utilizar un método híbrido, es decir, que trabaje con ambas fuentes

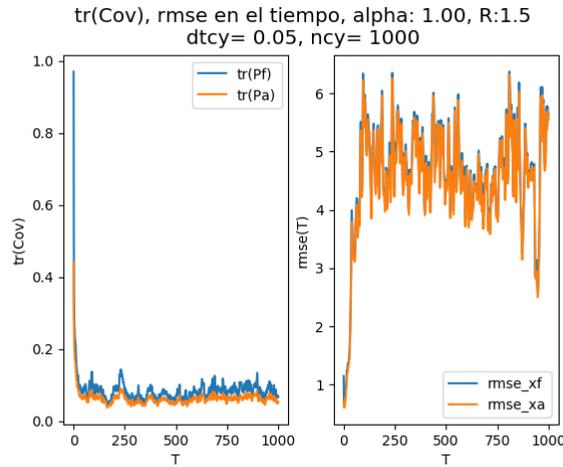


Figura 3.11: RMSE y trazas de matrices de covarianza en función del tiempo de asimilación a partir del modelo L96 sin factor de inflación. Experimento 3.5.

$\mathbf{R} = 1.5\mathbf{I}$		tr(Cov)	RMSE
Sin inflación	xf	0.097	4.772
	xa	0.078	4.682
Con inflación (1.36)	xf	0.458	0.672
	xa	0.391	0.593

Tabla 3.2: Promedio a lo largo de los ciclos de asimilación de las trazas de las matrices de covarianza y del RMSE a partir del modelo L96 en marco de modelo imperfecto, con y sin inflación. Experimentos 3.5 y 3.6.

de error a la vez.

Como resumen de esta sección: se mostraron resultados de experimentos con los modelos Lorenz 63 y Lorenz 96 en los que, al trabajar en presencia de error de modelo, el filtro de manera pura subestima la covarianza, lo cual genera que el filtro colapse y diverja. Se mostró luego como con la incorporación de un factor de inflación adecuado se puede representar el error de modelo, y así lograr la no divergencia y que la asimilación lograda sea un mejor estimador del estado real del sistema. Se concluye así la relevancia de la incorporación de estos parámetros multiplicativos de las covarianzas de error en los sistemas de asimilación.

Por otro lado, como motivación adicional para estimar de manera óptima el factor de inflación, se remarca que a medida que se trabaje en dimensiones más grandes, la subestimación de la covarianza del ensamble, además de estar producida por el error de modelo, también va a estar generada por el error de muestreo. De hecho, en problemas de alta dimensionalidad, es inevitable que

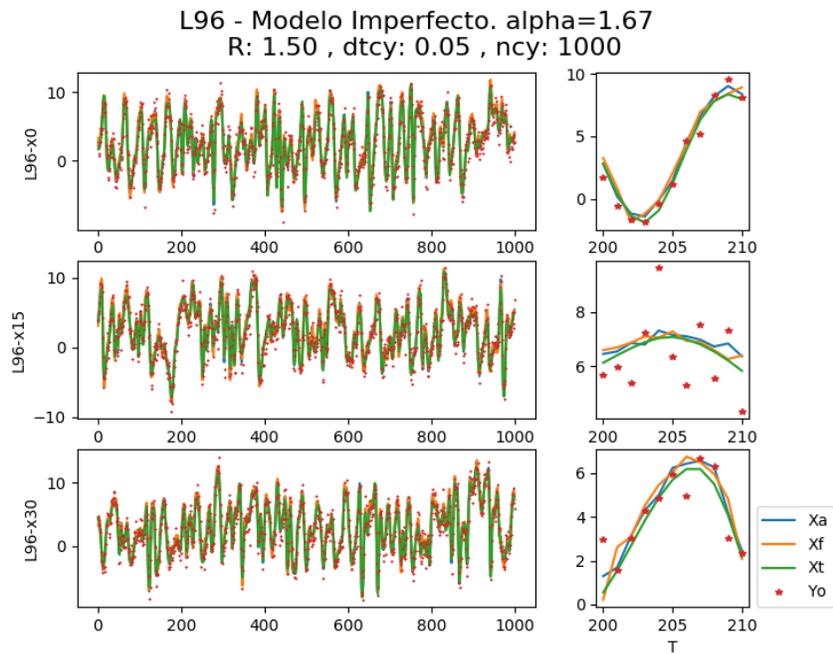


Figura 3.12: Series de tiempo y proceso de asimilación a partir del modelo L96 en marco de modelo imperfecto con factor de inflación. Experimento 3.6.

se cometa este error. Una de las maneras de subsanar este problema es con localización (como se mencionó en la introducción), pero otra manera también, es el uso de factores de inflación, que va a estar representando entonces no solo el error de modelo, sino también el error de muestreo.

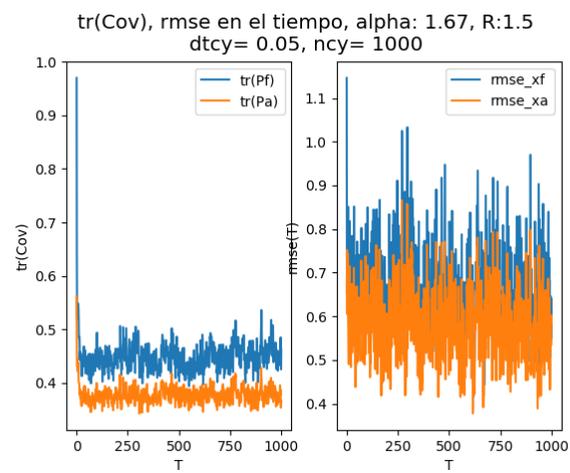


Figura 3.13: RMSE y trazas de matrices de covarianza en función del tiempo de asimilación a partir del modelo L96 con factor de inflación. Experimento 3.6.

Capítulo 4

Estimación por máximo-verosimilitud aproximada

En el capítulo anterior se demostró como en sistemas con error de modelo o con relativamente alta dimensionalidad se requiere mejorar la representación de la matriz de covarianza de las predicciones. Para desarrollar un método de estimación de parámetros de una representación de la matriz de covarianza se propone la utilización de la función verosimilitud en un intervalo relativamente largo de observaciones que denominaremos batch de observaciones.

Uno de los desafíos que existen en este tipo de problemas es que no conocemos en forma exacta la función de verosimilitud (i.e. intractable likelihood function). Para subsanar este inconveniente se propone trabajar con una verosimilitud aproximada asumiendo Gaussianidad, la log-verosimilitud aproximada, como se desarrolla en la Subsección 2.3.1, la aproximación resultante bajo esta hipótesis es ec. (2.30):

$$l(\theta) \approx -\frac{1}{2} \sum_{k=1}^K \left[(\mathbf{y}_k - \mathbf{H}\mathbf{x}_k^f)^T (\mathbf{H}\mathbf{P}_k^f \mathbf{H}^T + R)^{-1} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k) + \ln(|\mathbf{H}\mathbf{P}_k^f \mathbf{H}^T + R|) \right] + C$$

La metodología a seguir consiste en proponer una parametrización de la matriz de covarianza de la predicción y buscar los estimadores que maximizan la log-verosimilitud aproximada.

Esta aproximación es estocástica, ya que depende de la muestra que se tiene para realizar su estimación. Una realización distinta de los miembros del ensamble del pronóstico, en los distintos tiempos, dará una estimación de la log-verosimilitud aproximada diferente. Por ende, al ser estocástica, no es factible asegurar un valor óptimo sin depender de la muestra en el caso de muestras finitas.

Por otro lado, además de la estocasticidad, es necesario tener en cuenta que los valores determinados por los métodos de optimización podrían ser puntos extremos locales de la función de verosimilitud.

Para la determinación de los parámetros que maximizan la función verosimilitud aproximada, considerando la variabilidad resultante de la estocasticidad, se utiliza en este capítulo dos métodos: el método Grid Search y el algoritmo Nelder-Mead.

El método Grid Search permite obtener un estimador relativamente robusto. Este es en esencia un método de evaluación de la medida (la función verosimilitud) en experimentos para distintos hiperparámetros ligados a la matriz de covarianza, en el cual las variaciones de los hiperparámetros están representadas generalmente en una malla regular, asumiendo que no se conoce a priori la geometría de la función verosimilitud. Este es un método caro, y que es factible en el caso de unos pocos parámetros (hasta 5 sería razonable).

El algoritmo Nelder-Mead es un método numérico para minimizar una función objetivo en un espacio multidimensional que no requiere del gradiente, optimiza basado solo en los valores de la función objetivo (Subsección 2.3.2). Es independiente de una malla prefijada, pero si es dependiente de un valor inicial para realizar las iteraciones del algoritmo. A priori con este método no se conoce si el punto óptimo es un extremo global, local, o de inflexión. Sin embargo, es más viable que el Grid Search para escalar la técnica a más dimensiones (por ejemplo, se podría pensar en optimizar en el orden de hasta 20-50 parámetros).

En este capítulo se abordan tres problemas de estimación, mediante los métodos Grid Search y Nelder-Mead, de distintos parámetros asociados a la covarianza de la predicción (Subsección 2.2.8):

- Estimación del parámetro multiplicativo de la matriz de covarianza muestral de la predicción: α . Con la estimación de este parámetro se busca la estimación de la inflación multiplicativa óptima.
- Estimación del parámetro multiplicativo de la matriz de error de modelo: β . Con la estimación de este parámetro se busca la estimación de la inflación aditiva óptima.
- Estimación de ambos parámetros: (α, β) . Se busca la estimación de la combinación óptima del error multiplicativo y aditivo a la vez.

En este capítulo primero se trabaja en un marco perfecto para estudiar la asertividad de la técnica propuesta y la influencia del tamaño muestral. Luego, fijando un tamaño muestral y conociendo la técnica, se la aplicará en un marco imperfecto para abordar los tres problemas de interés. A base de esto, se sigue el siguiente esquema:

1. Asertividad de la técnica: Dado un tamaño de muestra, y para diferentes errores de observación, se trabaja en un marco de experimento perfecto en el cual se conoce el valor real de uno de los parámetros. Luego se realizan estimaciones del mismo utilizando los métodos Grid Search

y Nelder-Mead. Al conocer el valor real, este experimento permite validar estos métodos de búsqueda de parámetros óptimos, en el contexto de interés. Esto se aborda en la Subsección 4.1.1.

2. Influencia del tamaño muestral: Dado varios tamaños de muestra, y para un error de observación fijo, se trabaja en un marco de experimento perfecto. Se analiza como varía la estimación de uno de los parámetros variando el tamaño de la muestra. Al ser un marco de modelo perfecto la fuente de error proviene solo de una subrepresentación de la covarianza por error de muestreo, y por ende este es el error que el parámetro de la covarianza debe representar. Es esperable que aumentando el tamaño de la muestra, disminuya el requerimiento de una inflación aditiva o multiplicativa de la covarianza muestral, este experimento abordará estas sensibilidades de los parámetros. Esto se desarrolla en la Subsección 4.1.2.
3. Estimación de los parámetros: En el marco de un experimento imperfecto, dado un tamaño de muestra y para diferentes errores de observación, se trabaja con evaluaciones en grilla de la log-verosimilitud aproximada para estudiar la geometría del problema y analizar la presencia de máximos locales/globales. Luego se realizan estimaciones de los parámetros utilizando el método de Grid Search y el algoritmo Nelder-Mead. Esto se desarrolla en la Sección 4.2.

4.1. Análisis en el marco de modelo perfecto

4.1.1. Validación de la técnica

El objetivo de esta subsección es poder validar que la maximización de la log-verosimilitud aproximada por medio de los métodos Grid Search y Nelder-Mead, es una técnica válida para la estimación de parámetros de la covarianza en el marco de asimilación de datos. Se evalúa la estimación del parámetro aditivo β en experimentos controlados con el modelo L96.

Para validar esto, se diseña el siguiente marco experimental: se genera la verdad con un β true (β_t) conocido y luego se busca estimar β óptimo para el proceso de asimilación que se basa en asimilar con modelo perfecto. Se utilizan 1000 miembros ($N_e=1000$), así el parámetro no debe representar el error producido por la subestimación por tener pocas muestras. Si la metodología es adecuada, la estimación de β debería converger a estimaciones con valores cercanos a β_t .

Se realizaron dos experimentos con este diseño, en el primero se computa la log-verosimilitud aproximada para diferentes valores de β , se estudia la geometría del problema y se encuentran valores óptimos por medio de Grid Search. En el segundo, se optimiza la log-verosimilitud con el método Nelder-Mead. En ambos se espera recuperar el valor β true con el que se generó la verdad.

Experimento 4.1.1-a: Evaluación de la log-verosimilitud aproximada en función de β con L96 en marco perfecto.

Se realizaron varios experimentos de asimilación del tipo L63-per-infl-add (Sec. 3.2.2). Se proponen distintos valores de β entre 1 a 2, de a 0.01 y se conduce la asimilación. El objetivo es estimar el valor de β óptimo a través las dependencias de la verosimilitud en una grilla, es decir basados en el método de Grid Search.

Se asumió $\mathbf{Q}_f = 0.01\mathbf{I}$ y $\beta_t = 1.3$. Este proceso se realizó tres veces, usando diferentes valores de matriz de error de observación. Se utilizó $\mathbf{R} = r\mathbf{I}$, con $r \in \{0.5, 1, 1.5\}$, se utilizaron 1000 miembros de ensamble ($N_e=1000$) y 1000 observaciones ($K = 1000$). Estos experimentos recorren los diferentes valores para β , y para cada uno realiza un experimento L96-per-infl-add, en el que, luego de realizarse, se calculó el valor de log-verosimilitud aproximada, utilizando la ecuación (2.30), y el error cuadrático medio (RMSE).

Se muestra en la Figura 4.1 los valores de log-verosimilitud obtenidos como resultado de los experimentos. Se muestra que para cada matriz de error de observación asumida, la log-verosimilitud es cóncava, y alcanza valores máximos cercanos a 1.30. Se encuentra como a medida que el error de observación aumenta, la estimación de la función de log-verosimilitud toma valores más bajos, es decir, las observaciones son menos verosímil, lo cual es esperable.

Los valores β de grilla para los cuales se realiza el máximo de la verosimilitud y el mínimo de la función de error se encuentran en la Tabla 4.1 junto a los resultados del experimento 4.1.1-b.

Experimento 4.1.1-b: Optimización de la log-verosimilitud aproximada en función de β con L96 en marco perfecto.

Se analiza también los resultados obtenidos por medio del método Nelder-Mead. En el experimento 4.1.1-a se computó la log-verosimilitud aproximada en función de β , para diferentes valores de β . Se define entonces la función $g(\beta)$, que es la estimación de la log-verosimilitud aproximada dado un valor de β .

La función g depende de los otros parámetros del experimento que quedaron fijos: cantidad de observaciones ($K = 1000$), cantidad de miembros del ensamble ($N_e=1000$), y matrices de error de modelo y observaciones ($\mathbf{Q}_f = 0.05\mathbf{I}$ y $\mathbf{R} = r\mathbf{I}$). Para cada error de observación ($r \in \{0.5, 1, 1.5\}$) se tiene una función distinta.

Por medio del Algoritmo Nelder-Mead se busca el valor de β que minimiza $f(\beta) = -g(\beta)$. Este algoritmo de optimización no depende de una grilla fija predefinida, como es el caso del método Grid Search (detalles del algoritmo en la subsecc. 2.3.2).

En la Tabla 4.1 se muestran los resultados de este experimento y del anterior, es decir, los valores óptimos de β encontrados por medio del método Grid Search y Nelder-Mead (este último redondeado a 3 cifras decimales) para diferentes errores de observación asumidos. Todos estos resultados son estimaciones de $\beta_t = 1.3$.

Se muestra similitud en los resultados obtenidos por ambos métodos, y se muestra que las estimaciones obtenidas son cercanas al valor real. Se muestra

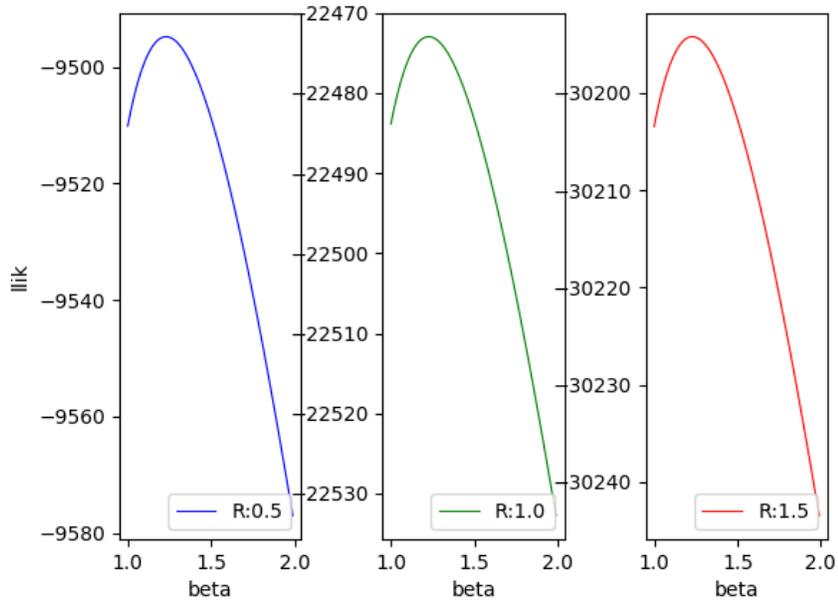


Figura 4.1: Log-verosimilitud aproximada en función del factor de inflación β para diferentes errores de observación usando Lorenz 96. Realizado en el marco de un experimento perfecto con $\beta_t = 1,3$. Experimento 4.1.1-a.

		r	Grid Search		Nelder-Mead
			Maximiza llik	Minimiza RMSE	
L96	β	0.5	1.23	1.36	1.230
		1.0	1.23	1.36	1.227
		1.5	1.23	1.37	1.229

Tabla 4.1: Estimaciones del parámetro multiplicativo del error de modelo $\beta_t = 1,3$ en el marco de modelo perfecto, por medio de los métodos Grid Search y Nelder-Mead, con el modelo L96. Experimentos 4.1.1-a y 4.1.1-b.

además los valores óptimos que minimizan el RMSE, y se puede ver que ambos resultados (máxima verosimilitud, o mínimo RMSE) distan en el mismo orden del valor real $\beta_t = 1.3$. Esto valida la estimación de máxima verosimilitud (aproximada) obtenida. En el caso de la estimación por mínimo RMSE, ésta solo puede calcularse en experimentos sintéticos, en los cuales la verdad es conocida.

Estos resultados dan indicios de que la técnica de maximizar la log-verosimilitud aproximada, ya sea por el método Grid Search como o por Nelder-Mead, para la estimación de parámetros de covarianza, converge a valores razonables del parámetro β . En la sección 4.2. se prosigue con aplicarla en un marco de experimento imperfecto, que es el marco de interés en la aplicación.

4.1.2. Influencia del tamaño muestral

En esta subsección se analiza la influencia del tamaño de la muestra en la estimación del parámetro α . Se interpreta la cantidad de miembros del ensamble como la cantidad de elementos de la muestra, ya que cada uno es una realización distinta del estado del sistema. Por ende, a medida que se tiene menos miembros de un ensamble, el error producido por error de muestreo aumenta.

Un tamaño de ensamble pequeño conlleva a una subestimación de la covarianza de la predicción muestral. La estimación de α o β buscan representar esta subestimación para aliviar este problema. Por ende, los parámetros óptimos de α y/o β , van a depender del tamaño del ensamble. Analizamos el caso de estimación solo de α , asumiendo que no se agrega error de modelo aditivo, es decir, asumiendo $\beta = 0$. Para estos experimentos, en el marco de experimento perfecto (sin error de modelo), las subestimaciones de la covarianza de predicción muestral van a deberse solo al error por tener pocas muestras en el ensamble. Este es el error que debe representar el factor de inflación, y se espera que a medida que la cantidad de miembros de ensamble aumente, los valores óptimos del factor de inflación disminuyan.

Experimento 4.1.2: Estimación de α con error de muestreo usando el modelo L96.

Se realizó un experimento que consiste en realizar muchos experimentos del tipo L96-per-infl-mult (Sec. 3.2.2). La diferencia entre ellos está en el factor de inflación α que se utiliza para amplificar la matriz de covarianza de los miembros del ensamble. Los factores de inflación probados van desde 1 a 2, con diferencia de 0.01.

Para evaluar las dependencias del factor de inflación con el tamaño del ensamble se utilizaron $N_e \in \{50, 100, 1000\}$ miembros del ensamble. No se utilizó ningún tipo de restricción sobre la covarianza (es decir, no se asumió ninguna hipótesis de localización.)

Estos experimentos recorren los diferentes valores de α , y para cada uno se realizó un experimento L96-per-infl-mult. Luego de realizarse, se calcula el valor de log-verosimilitud aproximada, utilizando la ec. (2.30), y el error cuadrático medio (RMSE). En todos los experimentos se asumió la matriz de covarianza de error de observación $\mathbf{R} = 1.5\mathbf{I}$ fija y se asumen 1000 observaciones. Además,

se asumió la matriz de error de modelo \mathbf{Q} nula.

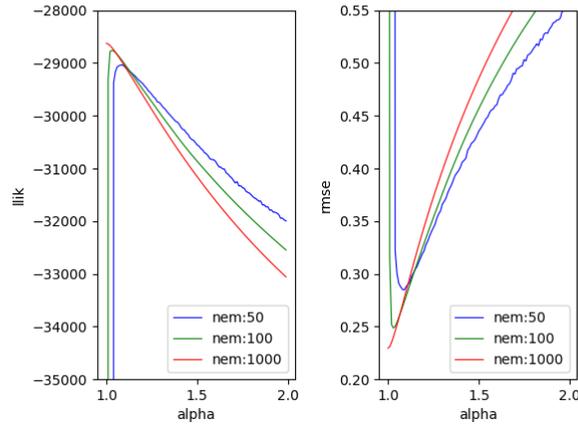


Figura 4.2: Función log-verosimilitud y RMSE en función del factor de inflación α para diferentes cantidades de miembros del ensamble usando Lorenz 96 en el marco de un modelo perfecto. Experimento 4.1.2.

Se muestran en la Figura 4.2 las funciones de log-verosimilitud y RMSE en función del factor de inflación α para los diferentes tamaños muestrales. Se muestra que al aumentar la cantidad de miembros del ensamble, disminuye el factor α óptimo. Además, en valores cercanos al óptimo, se observa que a medida que aumentan los miembros del ensamble, la log-verosimilitud aumenta y el RMSE disminuye.

Se muestra además que la log-verosimilitud aproximada sigue un comportamiento cóncavo y el RMSE convexo, lo que nos indica que, fijado el error de observación, habría un valor de factor de inflación óptimo que maximiza la log-verosimilitud, y otro que minimice el RMSE.

Además, se obtuvo que a mayor cantidad de muestras, está más marcada la concavidad de la log-verosimilitud aproximada y la convexidad del RMSE. Esto da noción de la incerteza para la estimación del parámetro, y de la sensibilidad del sistema, infiriendo que a mayor cantidad de miembros del ensamble, menor es la incerteza de la estimación y menor es la sensibilidad del sistema.

Se muestran en la Tabla 4.2 las estimaciones de α obtenidas al maximizar la log-verosimilitud aproximada con ambos métodos. En dichos resultados se muestra que, en este marco de modelo perfecto, con estado real de dimensión 40 (y por ende matrices de covarianza de dimensión 40x40), con 1000 muestras la estimación del factor de inflación es prácticamente 1. En esta dimensión de estado, para este tamaño muestral, se puede despreciar el error producido por el error de muestreo, es decir, la cantidad de muestras es suficiente. Se muestra además que se valida la hipótesis de que al disminuir la cantidad de miembros

		N_e	Grid Search		Nelder-Mead
			Maximiza llik	Minimiza RMSE	
L96	α	1000	1.00	1.00	1.001
		100	1.03	1.03	1.030
		50	1.08	1.09	1.087

Tabla 4.2: Estimación del factor de inflación α en el marco de modelo perfecto para diferentes tamaños muestrales, usando los métodos Grid Search y Nelder-Mead, con el modelo L96. Experimento 4.1.2.

del ensamble, la estimación del factor de inflación aumenta. Esto se debe a que el factor de inflación busca compensar el error producido por la subrepresentación de la covarianza muestral. Con este experimento se deja evidencia de que la estimación depende del tamaño muestral.

4.2. Estimación de parámetros en el marco de Modelo Imperfecto

En el marco de modelo imperfecto, se realizan diversos experimentos sintéticos. Dado un conjunto de observaciones y un modelo subrogado, se varían los parámetros de tal manera de estimar la covarianza de predicción que máxima la verosimilitud. Estos experimentos son similares a los de la sección anterior, pero aquí tenemos la presencia de dos fuentes de subestimación de la covarianza de predicción: por un lado, el error muestral y, por otro lado, el error de modelo. Ambas fuentes conllevan una subestimación de la covarianza de predicción. El valor verdadero de los parámetros, si es que existe, no se conoce.

En las siguientes tres subsecciones se trabaja con la variación de la log-verosimilitud aproximada en función de los parámetros α , β y (α, β) respectivamente. Se valúa esta función en una grilla y se grafica los resultados, para poder estudiar la geometría del problema. Y luego, se realizan estimaciones de los parámetros óptimos por medio de los métodos Grid Search y Nelder-Mead.

Se utiliza el modelo Lorenz 96 en todos los casos, y en la estimación de α además se utiliza también el modelo Lorenz 63. Se mantiene fija la cantidad de miembros del ensamble y observaciones, siendo estas últimas $K = 1000$ en todos los experimentos, y se varía el error de observación, utilizando 3 valores distintos ($r\mathbf{I}$, con $r = 0.5, 1, 1.5$). Para la estimación de α y (α, β) se utilizan ensambles de 100 miembros, y para la estimación de β se utilizan ensambles de 1000 miembros. Las grillas utilizadas para los parámetros α y β son las mismas que en los experimentos 4.1.1 y 4.1.2: varía entre 1 y 2 de a 0.01.

4.2.1. Estimación del factor de inflación α

Experimento 4.2.1-a: Optimización de la log-verosimilitud aproximada en función de α con L63 en un marco imperfecto.

Se realizó una serie de experimentos en los cuales en cada experimento, del tipo L63-imper-infl-mult (Sec. 3.2.2), se utiliza un factor de inflación α diferente, variando desde 1 a 2, con un salto de 0.01. Se realizaron tres series de estos experimentos para distintas covarianzas de observación $\mathbf{R} = r\mathbf{I}$, con $r \in \{0.5, 1, 1.5\}$. En cada experimento se calcula la log-verosimilitud aproximada y el error cuadrático medio (RMSE). No se representa error de modelo explícitamente, utilizando $\mathbf{Q} = 0$.

La Figura 4.3 muestra los valores obtenidos de log-verosimilitud aproximada (primer subfigura) y de RMSE (segunda subfigura) en función del factor de inflación α , para cada matriz de error de observación.

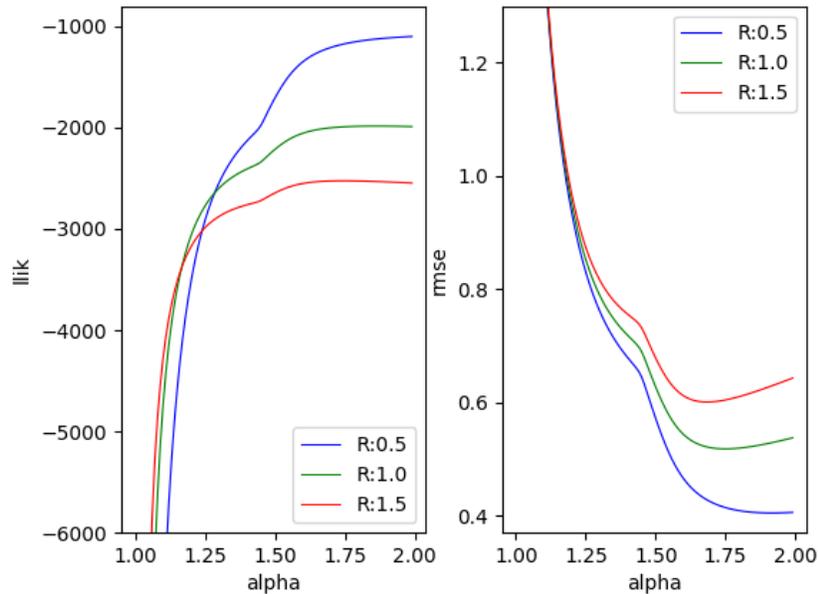


Figura 4.3: Log-verosimilitud aproximada y RMSE en función del factor de inflación α para diferentes errores de observación usando Lorenz 63, realizado en el marco de un experimento imperfecto. Experimento 4.2.1-a

Como es esperable, se observa que mientras menor es el error de las observaciones, menor es el error de la estimación. Por ende el filtro será un mejor estimador de la verdad, y consecuentemente, más verosímil. Esto último se puede observar al ver como la función de log-verosimilitud aproximada toma valores más altos, si el error de observación es menor.

Se muestra además que la log-verosimilitud sigue un comportamiento cón-

cavo, y el RMSE convexo, lo que nos indica que, fijado el error de observación, habría un valor de factor de inflación que maximice la log-verosimilitud, y uno que minimice el RMSE.

Se obtuvo también que a mayor error de observación, más marcada es la concavidad de la log-verosimilitud aproximada y la convexidad del RMSE. Como se ha mencionado en la subsección anterior, esto está ligado a la incerteza de la estimación del parámetro, y a la sensibilidad del sistema. Se obtiene como resultado en este caso que a mayor error de observación, menor es la incerteza de la estimación y menor es la sensibilidad del sistema, ya que el mínimo y el máximo de las respectivas funciones resulta mejor definido.

Usando la misma configuración en los experimentos se utiliza el algoritmo Nelder-Mead para determinar el α que optimiza la verosimilitud. Se trabaja con la log-verosimilitud aproximada como función del parámetro α , la llamamos $g(\alpha)$, y se optimiza con Nelder-Mead la función $f(\alpha) = -g(\alpha)$. Para cada error de observación fijado, se tiene una función distinta a optimizar.

Los valores de factor de inflación α en donde se realiza el máximo de la verosimilitud y el mínimo de la función de error con el método Grid Search, y los valores óptimos encontrados con Nelder-Mead, para este experimento, 4.2.1-a, y para el siguiente, el 4.2.1-b, se encuentran en la Tabla 4.3.

Experimento 4.2.1-b: Optimización de la log-verosimilitud aproximada en función de α con L96 en un marco imperfecto.

Este experimento es análogo al anterior, pero utilizando el Modelo Lorenz 96. Se realizó tres veces, asumiendo diferentes matrices de error de observación. En todos los casos se asumió matriz de error de modelo \mathbf{Q} nula.

Se muestra en la Figura 4.4, para cada error de observación, los valores de log-verosimilitud (primer subfigura) y RMSE (segunda subfigura), en función de los diferentes valores para el factor de inflación α , obtenidos como resultado en este experimento.

Se muestra, al igual que en la Figura 4.3, que mientras menor es el error de las observaciones, menor es el RMSE, y mayor es la log-verosimilitud. Del mismo modo, también se obtuvo que a mayor error de observación, más marcada es la concavidad de la log-verosimilitud y la convexidad del RMSE, y por ende el mínimo y el máximo de las respectivas funciones resulta mejor definido, lo cual implica que menor es la incerteza de la estimación y la sensibilidad del sistema.

Luego, de manera análoga al experimento anterior, se toma la log-verosimilitud aproximada en función del parámetro α como función $g(\alpha)$, y se optimiza con el algoritmo Nelder-Mead la función $f(\alpha) = -g(\alpha)$. Para cada error de observación fijado, se tiene una función distinta a optimizar.

Los valores en donde se realiza el máximo de la verosimilitud y el mínimo de la función de error con el método Grid Search, y los valores óptimos encontrados con Nelder-Mead, se encuentran en la Tabla 4.3.

Cabe recordar que la grilla utilizada tenía un máximo de 1.99, por lo que los valores encontrados por el método Grid-Search para $r = 0.5$ están diciendo que el α óptimo es mayor o igual a 2 para ambos modelos. Podría generarse una

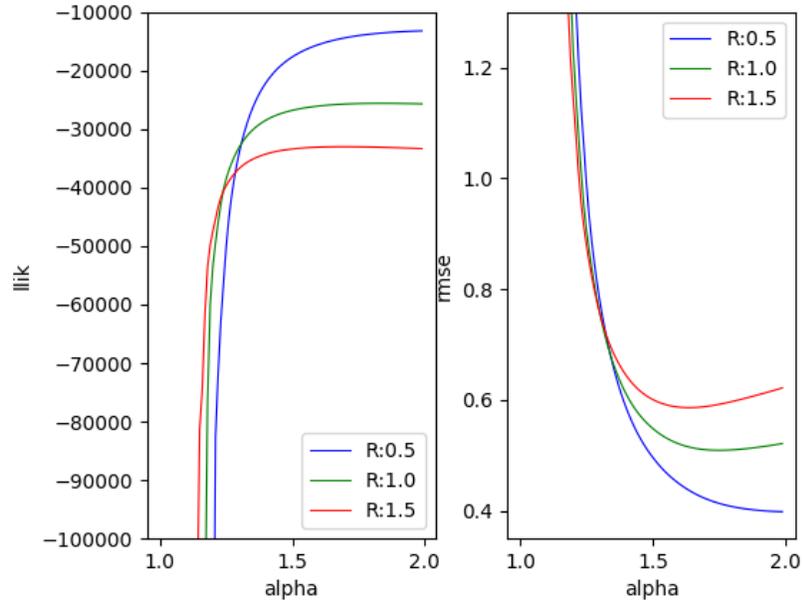


Figura 4.4: Log-verosimilitud aproximada y RMSE en función del factor de inflación α para diferentes errores de observación usando Lorenz 96, realizado en el marco de un experimento imperfecto. Experimento 4.2.1-b.

		r	Grid Search		Nelder-Mead
			Maximiza llik	Minimiza RMSE	
L63	α	0.5	1.99	1.92	2.219
		1.0	1.86	1.75	1.862
		1.5	1.74	1.68	1.744
L96	α	0.5	1.99	1.99	2.185
		1.0	1.84	1.75	1.836
		1.5	1.69	1.64	1.694

Tabla 4.3: Estimaciones del factor de inflación α en el marco de modelo imperfecto para distintos errores de observación, usando los métodos Grid Search y Nelder-Mead, con los modelos L63 y L96. Experimentos 4.2.1-a y 4.2.1-b.

segunda iteración para estos casos utilizando una grilla más amplia.

Para los otros casos de r , se muestra que los valores encontrados con el método Grid-Search son similares a los encontrados con Nelder-Mead. Por otro lado, contrario a lo que sucedía en modelo perfecto, aquí las estimaciones con máximo-verosimilitud dan parámetros levemente superiores a los dados por minimización del RMSE.

4.2.2. Estimación del parámetro de la covarianza β

Experimento 4.2.2: Optimización de la log-verosimilitud aproximada en función de β con L96 en un marco de modelo imperfecto.

Este experimento trabaja en el marco de modelo imperfecto utilizando Lorenz 96, similar al anterior, con la diferencia que se trabaja en optimizar las funciones de log-verosimilitud y RMSE en función del factor multiplicativo de la covarianza del error de modelo β , en lugar del factor de inflación α . Es decir, se realizan experimentos del tipo L96-imper-infl-add (Sec. 3.2.2).

Se muestran los resultados obtenidos en la Figura 4.5. Nuevamente se encuentra que mientras menor es el error de observación más verosímil es el filtro y menor la función de error. Las funciones de log-verosimilitud y RMSE son cóncava y convexa respectivamente, mostrando la existencia de un valor óptimo para el parámetro β en el rango definido. En particular se observa que es más marcada la concavidad de la log-verosimilitud cuando $\mathbf{R} = 0.5\mathbf{I}$, en cambio, en el RMSE no se observan diferencias en la convexidad en función del error de observación.

En la Tabla 4.4 se muestran los valores de β para los cuales se alcanza el máximo de la log-verosimilitud, y el mínimo del RMSE, con el método Grid Search, y los valores óptimos encontrados con Nelder-Mead, para cada valor de error de observación.

		r	Grid Search		Nelder-Mead
			Maximiza llik	Minimiza RMSE	
L96	β	0.5	0.07	0.07	0.072
		1.0	0.09	0.08	0.085
		1.5	0.09	0.09	0.094

Tabla 4.4: Estimaciones del parámetro multiplicativo de la covarianza β en el marco de modelo imperfecto para distintos errores de observación, usando los métodos Grid Search y Nelder-Mead, con el modelo L96. Experimento 4.2.2.

Los respectivos máximos y mínimos no siguen la misma lógica que con la estimación de α . En este caso, a medida que el error de observación es más grande, las estimaciones de β son mayores. Esto resulta contradictorio, y no queda claro en esta instancia cuál es la causa que hace que en este experimento con error aditivo de opuesto al realizado con el parámetro α , que maneja error

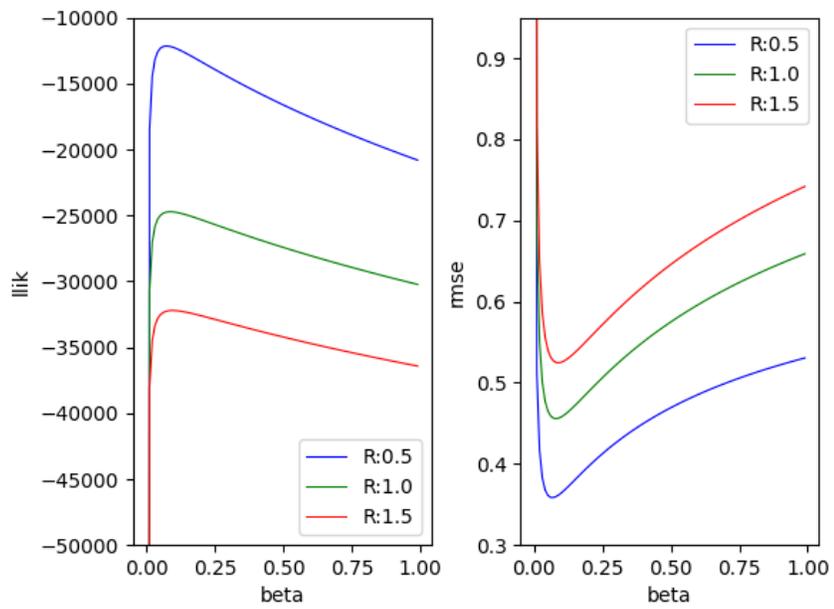


Figura 4.5: Log-verosimilitud aproximada y RMSE en función del factor de inflación β para diferentes errores de observación usando Lorenz 96, realizado en el marco de un experimento imperfecto. Experimento 4.2.2.

multiplicativo.

4.2.3. Estimación de ambas fuentes de error a la vez: (α, β) óptimo

Continuando con el marco de modelo imperfecto, se busca en este caso representar el error producido por asimilar con un modelo subrogado, por medio de amplificar las dos fuentes distintas con las que se trabajó hasta el momento: la covarianza del ensamble \mathbf{P}^f (utilizando un $\alpha \geq 1$), y la matriz de error de modelo \mathbf{Q} (utilizando $\beta \geq 0$).

Experimento 4.2.3: Optimización de la log-verosimilitud aproximada en función de (α, β) con L96 en un marco imperfecto.

Se realizaron la misma serie de experimentos pero considerando (α, β) como parámetros a determinar con el modelo Lorenz 96. Dado un conjunto de observaciones y un modelo subrogado, se varían los parámetros (α, β) de tal manera de estimar la covarianza de predicción que máxima la verosimilitud. Es decir, se realizan experimentos del tipo L96-imper-infl-mult-y-add (Sec. 3.2.2).

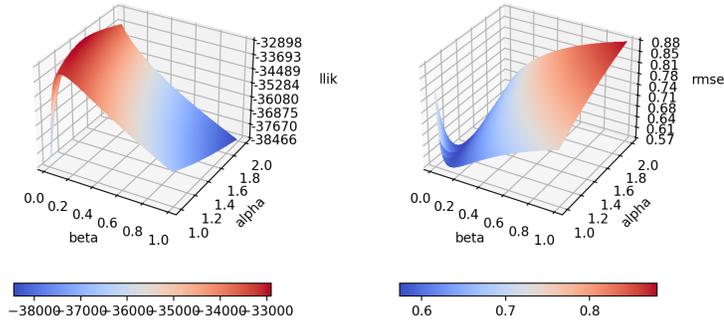


Figura 4.6: Log-verosimilitud aproximada y RMSE en función de (α, β) con $\mathbf{R} = 0.5\mathbf{I}$ usando Lorenz 96. Realizado en el marco de modelo imperfecto. Experimento 4.2.3.

Se muestra en la Figura 4.6 las representaciones de log-verosimilitud y RMSE generadas del experimento al evaluar la grilla descrita para $r = 0.5$. En la Figura no se grafica los valores de las funciones obtenidos para $\beta = 0$ ni $\beta = 0.01$, ya que dan valores más extremos, lo cual perjudicaba la gráfica.

Se observa que la función de log-verosimilitud aproximada y RMSE tienen máximo y mínimo globales respectivamente. Sin embargo, también se observa

que no son cóncava y convexas, y que esos puntos óptimos son poco identificables, ya que para valores de β pequeños, se observa un valle en las funciones.

Se muestra también que al variar ambos parámetros a la vez la log-verosimilitud aproximada resulta más sensible al parámetro β que al α , y que los valores bajos de β son los que resultan en valores mayores de la log-verosimilitud. Fijado un parámetro aditivo óptimo, se muestra que existe un parámetro multiplicativo óptimo, sin embargo, al ser menos sensible la log-verosimilitud a este y al tener un efecto de valle la función, se puede notar que podría haber, en vez de un valor óptimo, un set de valores óptimos que optimicen la log-verosimilitud.

En la Tabla 4.5 se observa los valores de (α, β) para los cuales se alcanza el máximo de la log-verosimilitud, y el mínimo del RMSE, con el método Grid Search, y los valores óptimos encontrados con Nelder-Mead para cada valor de error de observación. Se puede ver que en el caso de la Figura 4.6 el punto encontrado que maximiza esta estimación de la log-verosimilitud es (1.84,0.01).

También se observa que a medida que aumenta el error de observación, el par (α, β) que se estima tiene un valor de α menor y de β mayor, esto es coherente con lo que se observó con ambos parámetros por separado. Por otro lado, también se muestra que las estimaciones de valores óptimos encontradas por el algoritmo Nelder-Mead no son tan similares a las del Grid-Search, a excepción del caso de $r=1.5$. Las estimaciones encontradas por estos métodos fueron más similares cuando se optimizaron cada parámetro por separado.

		r	Grid Search		Nelder-Mead
			Maximiza llik	Minimiza RMSE	
L96	(α, β)	0.5	(1.84,0.01)	(1.70,0.01)	(1.913, 0.006)
		1.0	(1.63,0.01)	(1.45,0.02)	(1.590, 0.014)
		1.5	(1.46,0.02)	(1.34,0.03)	(1.462, 0.020)

Tabla 4.5: Estimaciones del par de parámetros (α, β) en el marco de modelo imperfecto para distintos errores de observación, usando los métodos Grid Search y Nelder-Mead, con el modelo L96. Experimento 4.2.3.

Conclusiones parciales:

Para cerrar, en este capítulo se obtuvo como resultados:

- Utilizar log-verosimilitud aproximada, aunque sea una aproximación, da una buena estimación de los parámetros (Sec. 4.1.1.). Dicha estimación dependen de la muestra, y del tamaño de la misma (Sec. 4.1.2.).
- El Algoritmo Nelder-Mead, en el marco de experimento imperfecto, converge, y converge a los máximos globales de la log-verosimilitud, valores para los que se obtuvo una estimación previa con el método Grid Search. (Sec. 4.2.)
- En presencia de error de modelo las estimaciones de α y β dependen del error de observación (Sec. 4.2.1 y 4.2.2). La relación de dependencia es

distinta para los dos parámetros, en el caso de α disminuye la estimación de cuando aumenta el error de observación, en el caso de β es a la inversa. Por otro lado, cuando no hay presencia de error de modelo, la estimación de β es independiente del error de observación (Sec. 4.1.1).

Con conclusión final, se obtiene que maximizando la log-verosimilitud aproximada, o Innovation Likelihood, es factible encontrar estimaciones de parámetros de la covarianza que logren representar el error producido por asimilar con modelos imperfectos. Esto se validó con los Modelos Lorenz 63 y Lorenz 96.

Capítulo 5

Estimación mediante el algoritmo expectation-maximization EM

Uno de los métodos que se propone utilizar para la estimación de parámetros ligados a la covarianza del pronóstico es el Expectation-Maximization, o EM. Se utilizarán las estimaciones realizadas en el capítulo anterior como referencia.

El interés de extender este algoritmo a la estimación de los parámetros de la covarianza, es que este método estima la log-verosimilitud explícita; es aplicable a la estimación de múltiples parámetros a la vez, siendo más escalable en la dimensión de los parámetros a estimar que el método Nelder-Mead (y por ende que Grid Search); y además fue probado para estimar tanto parámetros físicos como estadísticos, con sistemas caóticos no lineales, que maximicen la función de verosimilitud en contexto de observaciones parciales y variables escondidas.

Para esta tesina se utilizan dos variantes de este algoritmo. Una esta basada en la implementación clásica de la combinación EM-EnKS (subsec. 2.4.3) y otra, es un desarrollo nobel realizado a partir del EM-secuencial.

Para la estimación del parámetro β basta con la implementación clásica, ya que esta implementación en Dreano et al. (2016)[18] se utilizó para estimar la matriz de covarianza de error de modelo \mathbf{Q} , y el parámetro β es un factor multiplicativo de esta matriz. Por ende, este método puede modificarse levemente y utilizarse para estimar β en vez de \mathbf{Q} .

Para la estimación del factor de inflación α , el cual influye en los miembros del ensamble del pronóstico en cada paso de asimilación, la implementación clásica no es adecuada porque no permite la estimación de parámetros de la matriz de predicción. Para este caso, se desarrolla en esta tesina un algoritmo el cual denominamos el EM-secuencial. Este no requiere de un paso de suavizado y trabaja explícitamente en cada paso de asimilación, lo que da la posibilidad de incluir el factor de inflación α en las cuentas, y deducir una optimización del mismo.

En las siguientes dos subsecciones se derivan ambas teorías. Se muestran los experimentos desarrollados para la estimación de estos parámetros, y se contrasta con los resultados obtenidos en las secciones anteriores.

5.1. Estimación del parámetro de la covarianza β

5.1.1. EM y estimador de β

El objetivo es proponer una forma predefinida de la matriz de covarianza del error de modelo, y utilizar el algoritmo EM para encontrar el factor que maximice la función verosimilitud dado un conjunto de observaciones distribuidas en el tiempo. Llamamos \mathbf{Q}_f a la matriz de covarianza de base, la cual consideramos conocida, y \mathbf{Q} a la matriz de covarianza en el sistema de asimilación, que será de la forma $\mathbf{Q} = \beta \mathbf{Q}_f$, donde β es el factor que se quiere optimizar. La matriz \mathbf{Q}_f puede venir dada para el sistema en estudio por restricciones físicas, o ser tomada como la matriz de covarianza climática. En este trabajo solo hemos considerado formas diagonales de la \mathbf{Q}_f .

Para la implementación del estimador del factor β utilizando el algoritmo EM tradicional, nos basamos en el estimador $\hat{\mathbf{Q}}$ desarrollado para estimar la matriz de error de modelo \mathbf{Q} (ver Sección 2.4). A partir del estimador $\hat{\mathbf{Q}}$, generamos el estimador para el factor β , de la siguiente manera:

$$\hat{\beta}^{(i)} = \frac{\sum_{k=1}^n \sum_{j=1}^n (\hat{\mathbf{Q}}^{(i)}(\mathbf{Q}_f)^{-1})_{kj}}{n^2} \quad (5.1)$$

Se considera además conocida la matriz de covarianza del error de observaciones (\mathbf{R}), por lo que se reduce el algoritmo EM a la estimación única del parámetro multiplicativo β de la matriz \mathbf{Q}_f .

5.1.2. Implementación y experimentos

Basado en el Algoritmo 1 desarrollado en la subSección 2.4.3, se implementó esta modificación del algoritmo EM. Se agregó la deducción de $\hat{\beta}^{(i)}$ a partir de la estimación de $\hat{\mathbf{Q}}^{(i)}$.

Para los experimentos se asumió diagonal la matriz \mathbf{Q}_f , por lo que (5.2) queda reducida al promedio de los elementos de la diagonal, en lugar de tomar el promedio de todos los valores de la matriz $\hat{\mathbf{Q}}^{(i)}(\mathbf{Q}_f)^{-1}$. Además, se fuerza la búsqueda de matriz óptima \mathbf{Q} en el subespacio de las matrices diagonales, anulando los elementos por fuera de la diagonal antes de la próxima iteración. Esta hipótesis es similar a la utilizada en Miyoshi [9] para el factor de inflación. De esta manera, la estimación de β queda reducida a la siguiente ecuación, donde *diag* hace referencia a la proyección de la matriz al subespacio de las matrices diagonales:

$$\hat{\beta}^{(i)} = \frac{1}{N_x} tr(diag(\hat{\mathbf{Q}}^{(i)})(\mathbf{Q}_f)^{-1}) \quad (5.2)$$

Se muestra el pseudo-código en el Algoritmo 2, la implementación se llevó a cabo en python. En este Algoritmo se utilizó la misma notación que en el Algoritmo 1, y se llamó W al subespacio de las matrices de dimensión $N_x \times N_x$ diagonales.

Algorithm 2 EM-beta

Input: $\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, \mathcal{M}(\cdot), Q_f, Q_0$ and $\text{filt}(\cdot)$
 $\mathbf{x}_{0:K,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f \leftarrow \text{filt}(\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, Q_0)$ ▷ Filtro
 $\mathbf{x}_{0:K,1:N_e}^s \leftarrow \text{suav}(\mathbf{x}_{0,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f)$ ▷ Suavizado
for $i = 0, N_{iter}$ **do**
 $Q_{(i+1)} \leftarrow \frac{1}{k} \sum_{j=1}^k \Sigma_j^{(i)}$ ▷ Estimación de Q
 $Q_{(i+1)} \leftarrow \text{diag}(Q_{(i+1)})$ ▷ Proyección de Q a W
 $\beta_{(i+1)} \leftarrow \frac{1}{N_x} \text{tr}(Q_{(i+1)} * Q_f^{-1})$ ▷ Estimación de β
 $\mathbf{x}_{0:K,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f \leftarrow \text{filt}(\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, Q_j)$
 $\mathbf{x}_{0:K,1:N_e}^s \leftarrow \text{suav}(\mathbf{x}_{0:K,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f)$ ▷ Suavizado
end for
Output: $\beta_{1:N_{iter}}$

En estos experimentos se asume que en el sistema de asimilación posee un modelo simplificado del sistema real. En este contexto el error de modelo, y en particular su matriz de covarianza, deben representar las incertezas asociadas a estas simplificaciones del modelo.

Se genera un experimento con el modelo Lorenz 96 del tipo perfecto análogo al Experimento 4.1.1 con el fin de realizar una estimación conociendo el verdadero valor del parámetro. En cada uno de los experimentos que se realice, se comparará con las estimaciones obtenidas por medio de Grid Search y Nelder-Mead.

Experimento 5.1.0: Estimación de β con EM utilizando el modelo L96 en un marco Perfecto.

Este experimento es análogo al experimento 4.1.1 en el que se buscó estimar el parámetro β_t dentro de un contexto de experimento perfecto con Lorenz 96. La verdad fue generada asumiendo un error aditivo en cada paso por medio de la matriz $\beta_t \mathbf{Q}_f$, y se prueba diferentes procesos de asimilación, asumiendo como matriz de error de modelo $\beta \mathbf{Q}_f$.

Se comienza con un punto inicial, β_0 , y en cada iteración i se tiene un nuevo valor β_i dada por el algoritmo EM. El objetivo es que a medida que trascurren las iteraciones, las estimaciones converjan al valor real. Para cada valor β_i , se realiza de fondo un experimento del tipo L96-per-inf-add (sec. 3.2.2), para el cual se computa la log-verosimilitud estimada y el RMSE. El algoritmo EM se realizó dos veces, inicializándolo en diferentes puntos. Se utilizaron como puntos iniciales $\beta_0 = 1$ y $\beta_0 = 25$, es decir, se inicializó el algoritmo con $\mathbf{Q}_0 = 25 \mathbf{Q}_f$ y con $\mathbf{Q}_0 = \mathbf{Q}_f$. Los resultados fueron similares.

En la Figura 5.1 se muestra la estimación del parámetro β en cada iteración del algoritmo, y los valores de la log-verosimilitud aproximada y del RMSE para cada caso. Se observa que el algoritmo converge: se estaciona las estimaciones de β ; el valor de log-verosimilitud aproximada se observa como en tendencia a maximizarse, y, por otro lado, en cada iteración, se logra un valor menor del RMSE.

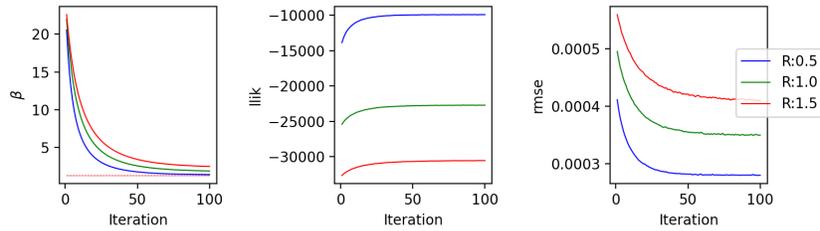


Figura 5.1: Experimento 5.1.0. Estimación de $\beta_t = 1.3$. Log-verosimilitud aproximada y RMSE, en función de cada iteración del algoritmo EM, para distintos errores de observación con un valor inicial del parámetro $\beta_0 = 25$, i.e. $Q_0 = 25Q_f$. Realizado en el marco de modelo perfecto.

Una vez que se estaciona el algoritmo, cada valor podría considerarse una estimación de β . Para cada error de observación, se computó el promedio de las últimas 20 iteraciones para considerarlo como estimador final del parámetro. Se muestra estos resultados para ambos puntos iniciales en la Tabla 5.1 junto con los resultados del Experimento 4.2.2.

		r	Grid Search		Nelder-Mead	EM	
			Max llik	Min RMSE		$\beta_0 = 1$	$\beta_0 = 25$
L96	β	0.5	1.23	1.36	1.230	1.374	1.445
		1.0	1.23	1.36	1.227	1.760	1.950
		1.5	1.23	1.37	1.229	2.266	2.573

Tabla 5.1: Estimación del parámetro multiplicativo del error de modelo $\beta_t = 1.3$ en el marco de modelo perfecto para distintos errores de observación, usando los métodos Grid Search, Nelder-Mead y EM, con el modelo L96. Experimento 5.1.0.

Se muestra que el algoritmo, para ambos puntos de inicio y para todos los errores de observación, converge, pero que sobreestima el valor real (1.3). Mientras mayor es el error de observación, mayor es la sobreestimación del valor del parámetro. Se prosigue con aplicarlo en el caso de modelo imperfecto, tanto con el modelo Lorenz 63 como Lorenz 96.

Experimento 5.1.1: Estimación de β con EM utilizando el modelo L63 en un marco Imperfecto.

Este experimento es análogo al experimento 4.2.2 (sec. 4.2.2), pero con el modelo Lorenz 63 y utiliza el algoritmo EM para estimar el parámetro β . Para cada valor de β , se realiza un experimento del tipo L63-imper-inf-add (sec. 3.2.2). Al igual que para el exp. 4.2.2. se asumió $\mathbf{Q} = \beta\mathbf{I}$.

Se realizan 50 iteraciones del algoritmo EM. Se utiliza el promedio de los últimos 10 valores como estimación del parámetro beta. Se muestran los resultados de este experimento en la Tabla 5.2, junto a las estimaciones realizadas con los métodos Grid Search y Nelder-Mead.

Experimento 5.1.2: Estimación de β con EM utilizando el modelo L96 en un marco Imperfecto.

Este experimento es análogo al Experimento 4.2.2 (sec. 4.2.2) pero utiliza el algoritmo EM para estimar el parámetro β . Se asumió $\mathbf{Q} = \beta\mathbf{I}$, y para cada valor de β se realiza un experimento del tipo L96-imper-infl-add (Sección 3.2.2).

Los resultados de convergencia de este experimento se muestran en Fig. 5.2 y en la Tabla 5.2.

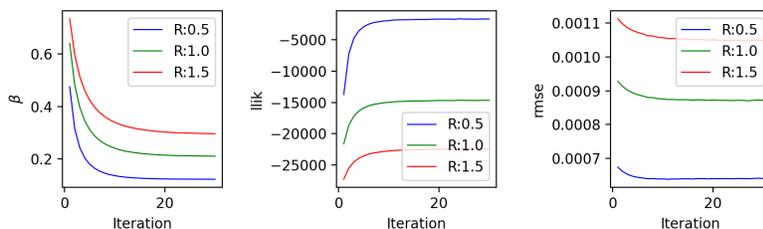


Figura 5.2: Estimación de β , log-verosimilitud aproximada y RMSE, con L96, en función de cada iteración del algoritmo EM, para distintos errores de observación. Realizado en el marco de un experimento imperfecto. Experimento 5.1.2.

Tomando como valor de referencia los resultados encontrados con el método Grid Search, se muestra que para el caso del modelo Lorenz 63 se subestiman los valores, mientras que con el modelo Lorenz 96 se sobrestima. Las diferencias que se observan son mayores, tanto en términos relativos como absolutos, que en las observadas en la Tabla 5.1. Además, se observa que a medida que el error de observación es más grande, el error por sobreestimación del parámetro β es mayor, resultado análogo al encontrado con los experimentos 5.1.0.

		r	Grid Search		Nelder-Mead	EM
			Max llik	Min RMSE		
L63	β	0.5	0.2	0.18	0.196	0.152
		1.0	0.24	0.22	0.238	0.186
		1.5	0.26	0.25	0.264	0.211
L96	β	0.5	0.07	0.07	0.072	0.124
		1.0	0.09	0.08	0.085	0.217
		1.5	0.09	0.09	0.094	0.306

Tabla 5.2: Estimación del parámetro multiplicativo del error de modelo β en el marco de modelo imperfecto para distintos errores de observación, usando los métodos Grid Search, Nelder-Mead y EM, con los modelos L63 y L96. Experimentos 5.1.1 y 5.1.2.

5.2. Estimación del factor de inflación α

Se busca en esta sección extender el método EM (Sec. 2.4) para estimar el parámetro α . Se asume que en cada tiempo de asimilación se inflarán los miembros del ensamble del pronóstico respecto a la media, y así la matriz de covarianza del ensamble para el tiempo k pasará a ser de P_k^f a αP_k^f . Esto se realizará antes de la asimilación de las observaciones.

Se busca manipular la verosimilitud completa de forma tal que aparezca la probabilidad de cada observación dependiendo de la observación/es anterior/es. De esta manera se espera que resulte explícita la presencia de la covarianza de la predicción en el paso de Expectation para luego poder maximizar la función intermedia resultante en función del parámetro α . Así, se utiliza la deducción del EM secuencial, que no utiliza un suavizador, en lugar del EM tradicional.

5.2.1. EM secuencial y estimador de α

La log-verosimilitud completa puede ser expresada en una forma secuencial usando probabilidades condicionales de la observación previa, como se realiza en [16].

$$p_\theta(\mathbf{y}_{1:K}) = \prod_{k=1}^K p_\theta(\mathbf{y}_k | \mathbf{y}_{1:k-1}) \quad (5.3)$$

Esta expresión puede ser naturalmente expresada en términos de la verosimilitud de las observaciones y la función de densidad de la predicción o pronóstico.

$$p_\theta(\mathbf{y}_{1:K}) = \prod_{k=1}^K \int p_\theta(\mathbf{y}_k | \mathbf{x}_k) p_\theta(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k \quad (5.4)$$

Se puede usar la expresión secuencial de la log-verosimilitud, (5.4), para derivar el algoritmo EM. Multiplicando y dividiendo por alguna función $q(\mathbf{x}_k)$ (cuyo soporte debe incluir al de la predicción $p_\theta(\mathbf{x}_k | \mathbf{y}_{1:k-1})$) se tiene

$$l(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:K}) = \log \prod_{k=1}^K \int q_{\boldsymbol{\theta}'}(\mathbf{x}_k) \frac{p_{\boldsymbol{\theta}}(\mathbf{y}_k|\mathbf{x}_k) p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{q_{\boldsymbol{\theta}'}(\mathbf{x}_k)} d\mathbf{x}_k \quad (5.5)$$

Usando la desigualdad de Jensen,

$$l(\boldsymbol{\theta}) \geq \sum_{k=1}^K \int q_{\boldsymbol{\theta}'}(\mathbf{x}_k) \log \left(\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_k|\mathbf{x}_k) p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{q_{\boldsymbol{\theta}'}(\mathbf{x}_k)} \right) d\mathbf{x}_k \quad (5.6)$$

Se define la función intermedia para el algoritmo EM como

$$G(q, \boldsymbol{\theta}) = \sum_{k=1}^K \int q_{\boldsymbol{\theta}'}(\mathbf{x}_k) \log \left(\frac{p_{\boldsymbol{\theta}}(\mathbf{y}_k|\mathbf{x}_k) p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{q_{\boldsymbol{\theta}'}(\mathbf{x}_k)} \right) d\mathbf{x}_k \quad (5.7)$$

La función intermedia depende de la función q y del parámetro $\boldsymbol{\theta}$. Para determinar la función q que maximiza G con $\boldsymbol{\theta}$ fijo, usamos Bayes:

$$p_{\boldsymbol{\theta}}(\mathbf{y}_k|\mathbf{x}_k) p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k}) p_{\boldsymbol{\theta}}(\mathbf{y}_k|\mathbf{y}_{1:k-1}) \quad (5.8)$$

Reemplazando (5.8) en (5.6), se deduce que la función q que maximiza la función intermedia es la densidad posterior $q_{\boldsymbol{\theta}'}^*(\mathbf{x}_k) = p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k})$ para $\boldsymbol{\theta} = \boldsymbol{\theta}'$, la cual, de hecho, satisface la igualdad. Con esto se tiene el paso E del algoritmo (*expectation step*). Para el paso de maximización, necesitamos encontrar el parámetro $\boldsymbol{\theta}$ que maximiza la función intermedia $G(q^*, \boldsymbol{\theta})$, es decir, $\boldsymbol{\theta}$ tal que

$$\begin{aligned} 0 &= \nabla_{\boldsymbol{\theta}} G(q^*, \boldsymbol{\theta}) \\ &= \sum_{k=1}^K \int q_{\boldsymbol{\theta}'}^*(\mathbf{x}_k) \nabla_{\boldsymbol{\theta}} \log (p_{\boldsymbol{\theta}}(\mathbf{y}_k|\mathbf{x}_k) p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k-1})) d\mathbf{x}_k \end{aligned} \quad (5.9)$$

Reemplazamos $q_{\boldsymbol{\theta}'}^*(\mathbf{x}_k)$ por $p_{\boldsymbol{\theta}'}(\mathbf{x}_k|\mathbf{y}_{1:k})$. Luego, en el caso de que $\boldsymbol{\theta}$ sea el factor de inflación α , esto solo afecta a la densidad de la predicción, ya que la log-verosimilitud observacional es independiente de $\boldsymbol{\theta}$, quedando la ecuación (5.9) de la siguiente manera:

$$\sum_{k=1}^K \int p_{\boldsymbol{\theta}'}(\mathbf{x}_k|\mathbf{y}_{1:k}) \nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k = 0 \quad (5.10)$$

La densidad de la predicción es asumida como Gaussiana, en la cual la covarianza ha sido multiplicada por el factor de inflación α ,

$$p_{\alpha}(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \frac{1}{[\det(\alpha \mathbf{P}_k^f)]^{1/2}} \exp \left[-\frac{1}{2} \left\| \mathbf{x}_k - \bar{\mathbf{x}}_k^f \right\|_{\alpha \mathbf{P}_k^f} \right] \quad (5.11)$$

donde denotamos $\left\| \mathbf{x}_k - \bar{\mathbf{x}}_k^f \right\|_{\alpha \mathbf{P}_k^f} \triangleq (\mathbf{x}_k - \bar{\mathbf{x}}_k^f)^{\top} (\alpha \mathbf{P}_k^f)^{-1} (\mathbf{x}_k - \bar{\mathbf{x}}_k^f)$.

Luego, $\det(\alpha \mathbf{P}_k^f) = \alpha^{N_x} \det(\mathbf{P}_k^f)$. Así, reemplazando (5.11) en (5.9), y despejando el factor de inflación α , el paso de maximización del algoritmo queda dado por

$$\alpha = \frac{1}{N_x K} \sum_{k=1}^K \int p_{\alpha'}(\mathbf{x}_k | \mathbf{y}_{1:k}) \left\| \mathbf{x}_k - \bar{\mathbf{x}}_k^f \right\|_{\mathbf{P}_k^f} d\mathbf{x}_k. \quad (5.12)$$

En el caso del filtro de Kalman por ensamble, en la i -ésima iteración del EM dado α_i , el paso E del algoritmo dado por la ec. (5.11) será estimado con el ensamble de análisis como muestra, y la función delta de Dirac δ , de la siguiente manera:

$$p_{\alpha_i}(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{1}{N_e} \sum_{j=1}^{N_e} \delta(\mathbf{x}_k^{a(j,i)} - \mathbf{x}_k) \quad (5.13)$$

donde $\mathbf{x}_k^{a(j,i)}$ son los miembros del ensamble de análisis luego de aplicar EnKF usando los miembros del pronóstico con el factor α_i .

Usando Monte Carlo, integrando (5.12), el paso de maximización es

$$\hat{\alpha}^{(i+1)} = \frac{1}{K N_x N_e} \sum_{k=1}^K \sum_{j=1}^{N_e} \left\| \mathbf{x}_k^{a(j,i)} - \bar{\mathbf{x}}_k^{f(i)} \right\|_{\hat{\mathbf{P}}_k^{f(i)}} \quad (5.14)$$

donde $\hat{\mathbf{P}}_k^{f(i)} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{x}_k^{f(j,i)} - \bar{\mathbf{x}}_k^f) (\mathbf{x}_k^{f(j,i)} - \bar{\mathbf{x}}_k^f)^\top$.

El estado del pronóstico usado para computar $\hat{\mathbf{P}}_k^{f(i)}$ está dado por

$$\mathbf{x}_k^{f(j,i)} = \mathcal{M}(\mathbf{x}_{k-1}^{a(j,i)}) \quad (5.15)$$

el cual se calcula con la evolución, a través del modelo, de los miembros de análisis del tiempo anterior ($k-1$) sin aplicar el factor de inflación en el tiempo k .

La expresión (5.14) captura las diferencias que se producen entre el ensamble de pronóstico y el ensamble de análisis, y cuantifica estas diferencias con la covarianza muestral de predicción. De hecho, en el caso perfecto que no se requiere un factor de inflación, el ensamble de análisis sería similar al del pronóstico, y por ende las diferencias computadas estarían estimando en cada tiempo algo similar a la covarianza $\hat{\mathbf{P}}_k^{f(i)}$, y la expresión (5.14) estimaría un valor de α cercano a 1.

En resumen, el algoritmo EM establece un método iterativo para encontrar el máximo de la log-verosimilitud secuencial. En el primer paso se realiza el proceso de asimilación con un valor de α_0 inicial, el cual se utiliza en cada uno de los ciclos de asimilación. Al finalizar, se calcula la estimación de α_1 como en (5.14), y así iterativamente.

5.2.2. Implementación y experimentos

Se implementó en Python el estimador iterativo descrito del factor de inflación α basado en el algoritmo EM. El parámetro a actualizar solo es α , se toma \mathbf{Q} y \mathbf{R} conocidos, y para esta implementación, se asume $\beta = 0$.

En cada iteración se calcula la ecuación (5.14) para estimar α_i , y con este factor se actualiza el filtro y se recalculan los ensambles de análisis y pronóstico, a partir de los cuales, se calculará la estimación de α_{i+1} , y así sucesivamente hasta que se realicen el número de iteraciones N_{iter} definido.

Se muestra en el Algoritmo 3 el pseudo-código del método propuesto. Notar que en cada iteración se está haciendo un filtrado como en el caso convencional, el único costo computacional extra es la estimación del α en el algoritmo.

Algorithm 3 EM-alpha

Input: $\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, \mathcal{M}(\cdot), Q, \text{filt}(\cdot)$ and α_0
for $i = 0, N_{iter}$ **do** ▷ Filtro
 $\mathbf{x}_{0:K,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f \leftarrow \text{filt}(\mathbf{x}_{0,1:N_e}^a, \mathbf{y}_0, Q, \alpha_i)$
 $\alpha_i \leftarrow \text{est-alpha}(\mathbf{x}_{0:K,1:N_e}^a, \mathbf{x}_{0:K,1:N_e}^f)$ ▷ Estimación de α_i (ec.(5.14))
end for
Output: $\alpha_{0:N_{iter}}$

Se realizaron experimentos con este algoritmo en el marco de modelo imperfecto usando L63 y L96, análogos a los Experimentos 4.2.1-a y 4.2.1-b del capítulo anterior.

Experimentos con error paramétrico

Experimento 5.2.1: Estimación de α con EM utilizando el modelo L63 en un marco Imperfecto.

En este experimento se deja fijo el error de observación, se asume $\mathbf{R} = 0.5\mathbf{I}$, y para cada valor de α se realiza un experimento del tipo L63-imper-infl-mult (Sección 3.2.2). Este es similar al Experimento 4.2.1-a, pero utiliza el algoritmo EM para estimar el parámetro α . Se recuerda con respecto a los parámetros que, en este marco de modelo imperfecto, se utilizan los parámetros $\sigma = 10$, $\beta = 8/3$ y $\rho = 28$, para generar la verdad, y los parámetros $\sigma = 11.5$, $\beta = 2.87$ y $\rho = 32$ para asimilar.

Se genera el estado real con el modelo L63 asumiendo $\Delta t = 0.01$, error aditivo nulo ($\mathbf{Q}_t = 0$). Se asumen 1000 observaciones sintéticas ($K = 1000$), se utilizan 100 miembros de ensamble ($N_e = 100$), se inicializa el algoritmo con un valor inicial $\alpha_0 = 5$, y se realizan 30 iteraciones del algoritmo EM.

Se muestran el resultado del experimento en la figura 5.3. Se observa que el algoritmo se estabiliza y converge a partir de aproximadamente la décima iteración. Se utiliza el promedio de los últimos 10 valores como estimación del parámetro α , y se muestran en la Tabla 5.3.

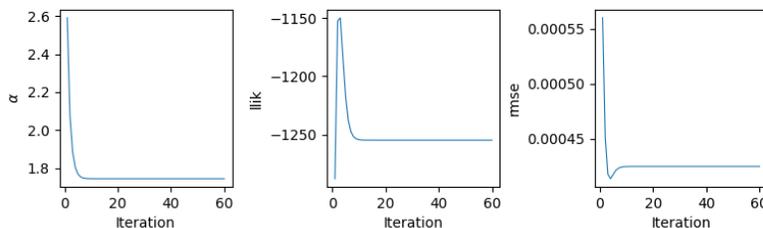


Figura 5.3: Experimento 5.2.1: Estimación de α con EM y modelo imperfecto usando L63.

Experimento 5.2.2: Estimación de α con EM utilizando el modelo L96 en un marco Imperfecto.

Este experimento es similar al anterior pero usando el modelo Lorenz 96. Se deja fijo el error de observación, se asume $\mathbf{R} = 0.5\mathbf{I}$, y para cada valor de α se realiza un experimento del tipo L96-imper-infl-mult (Sección 3.2.2). Por ende, en este marco de experimento con modelo imperfecto, se utiliza un parámetro de forzado $F = 8$ para generar la verdad, y $F = 10$ para realizar la asimilación.

Se genera un estado real con el modelo L96 asumiendo $\Delta t = 0.05$, error aditivo nulo ($\mathbf{Q}_t = 0$). Se asumen 1000 observaciones sintéticas ($K = 1000$), y se utilizan 100 miembros de ensamble ($N_e = 100$). Se inicializa el algoritmo con un valor inicial $\alpha_0 = 1$.

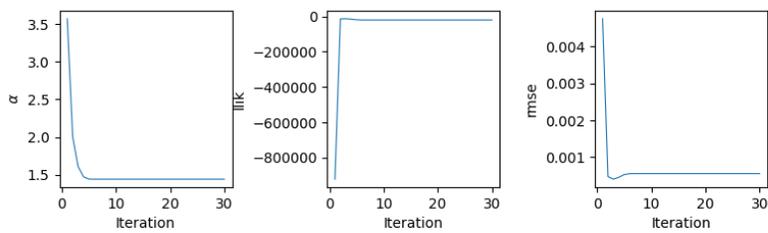
En figura 5.4 se observa que el algoritmo se estabiliza y converge a partir de aproximadamente la quinta iteración. Se utiliza el promedio de los últimos 10 valores como estimación del parámetro α , y se muestra en la Tabla 5.3).

		r	Grid Search		Nelder-Mead	EM
			Max llik	Min RMSE		
L63	α	0.5	1.99	1.92	2.219	1.744
L96	α	0.5	1.99	1.99	2.185	1.440

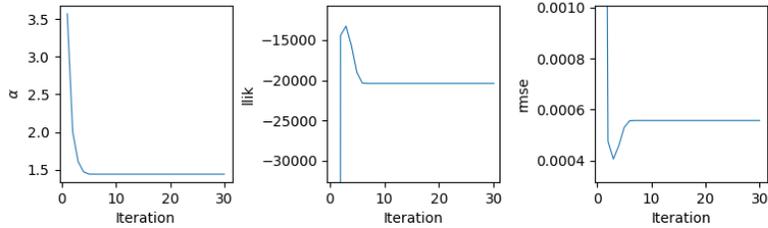
Tabla 5.3: Estimación del factor de inflación α en el marco de modelo imperfecto, usando los métodos Grid Search, Nelder-Mead y EM, con los modelos L63 y L96. Experimentos 5.2.1 y 5.2.2.

Como resultado de los experimentos en las figuras 5.3 y 5.4 se muestra que, a pesar de que las estimaciones de α son siempre descendientes, en las log-verosimilitudes se realiza un efecto de lomada y de manera análoga en el RMSE un efecto de valle.

Por otro lado, si se toma como referencia los valores óptimos encontrados con los métodos Grid Search y Nelder-Mead (cabe recordar que el método Grid



(a) Resultado del experimento



(b) Resultado del experimento con zoom

Figura 5.4: Experimento 5.2.2: Estimación de α con EM y modelo imperfecto usando L96.

Search se basó en una grilla con un mínimo de 1 y máximo de 2) se muestra que no hay convergencia en la técnica a los valores óptimos de RMSE y función de log-verosimilitud. El método EM subestima el valor óptimo del parámetro α .

Viendo la subestimación del algoritmo y los efectos de lomada y valle, podría interpretarse que en esos picos se encuentra el óptimo, pero que el algoritmo no pudo converger allí. Sin embargo, para los casos de comparación de la Tabla, en las iteraciones donde se realizan estos picos, el EM estima los valores $\hat{\alpha} = 1.887$ y $\hat{\alpha} = 1.606$ para L63 y L96 respectivamente, por lo cual sigue estando alejado de los valores óptimos según Nelder-Mead de 2.219 y 2.189.

Un punto a discutir, además de la subestimación, es por qué se observa este efecto de lomada. El Algoritmo EM en cada iteración busca maximizar la log-verosimilitud completa, y si bien la que se grafica la log-verosimilitud aproximada, se esperaría que este gráfico sea creciente. A continuación la técnica es evaluada en experimentos en diferentes marcos de modelos imperfectos. En cada caso se varía el error de modelo a representar, y se varían las variables del experimento. Se busca estudiar cuáles de estas variaciones influyen en la estimación/subestimación del valor del parámetro a estimar, y junto a esto, si los efectos de lomada y valle varían. Se busca así caracterizar el efecto y la interpretación sobre los procesos que lo producen.

Experimento 5.2.3: Estimación de α con EM utilizando L96 en diferentes grados de error paramétrico.

Este experimento es similar al 5.2.2, pero se va a realizar el experimento 5 veces, donde en cada caso se varía el error paramétrico que tiene el modelo de predicción de la asimilación. El modelo L96 utiliza el parámetro de forzado F . El parámetro clásico es $F = 8$, y el parámetro que se asumió para representar error paramétrico en experimentos anteriores es $F = 10$. Para el presente experimento, se utilizan los siguientes valores de F : 8, 8.1, 8.5, 9 y 10. Con $F = 10$ se reproduce el resultado obtenido en el experimento 5.2.2.

Se utiliza EM para la estimación de α con el modelo L96 en diferentes marcos de experimento imperfecto. Se realiza el experimento 4 veces, en cada una se usa un valor distinto de forzado F del modelo con el que se asimila: 8.1, 8.5, 9, 10. Con $F = 10$ se reproduce el resultado obtenido en el experimento 5.2.2.

En la Figura 5.5 y en la Tabla 5.4 se muestran los resultados obtenidos. En dicha tabla se muestra con $F = 10$ el valor del experimento anterior.

La Figura 5.5 muestra aun cuando α parece converger a valores óptimos para los distintos valores de error paramétrico. A medida que aumenta el error de modelo, la estimación del factor de inflación aumenta, y en ese contexto, la log-verosimilitud disminuye y el RMSE aumenta. Se muestra que cuando el modelo de asimilación se acerca al real (F se acerca a 8), las estimaciones de α se acercan a 1. Se muestra también que los efectos de lomada y de valle disminuyen al disminuir el error de modelo, siendo este efecto prácticamente nulo en los casos $F = 8.1$.

Se utiliza el promedio de los últimos 10 valores como estimación del parámetro α . Se muestran estas estimaciones en la Tabla 5.4. Se muestra que al

disminuir el error de modelo, el algoritmo termina estimando valores cercanos a 1, el cual (salvo error de muestreo) es el valor óptimo para el caso del experimento perfecto. Por ende, el método subestima los valores de α en contexto de experimento imperfecto, pero al disminuir el error de modelo, realiza buenas estimaciones.

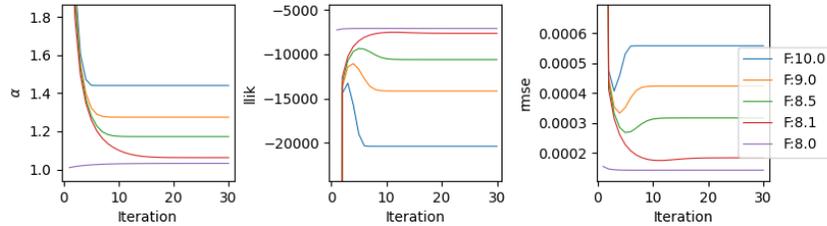


Figura 5.5: Experimento 5.2.3: Estimación de α en distintos contextos de error paramétrico, usando L96.

L96	
F	$\hat{\alpha}$
10	1.440
9	1.274
8.5	1.172
8.1	1.062
8.0	1.032

Tabla 5.4: Estimación del factor de inflación α en diferentes marcos de modelos imperfectos, usando el algoritmo EM el modelo L96. Experimento 5.2.3.

Experimento 5.2.4: Variando los parámetros de los experimentos imperfectos en la estimación de α con EM.

Se realiza una serie de experimentos, con los que se busca variar los parámetros involucrados en el diseño de los experimentos anteriores: valor de factor de inflación inicial, número de miembros del ensamble, y número de observaciones. Esto se realiza con dos objetivos: uno, es entender si estas variables influyen el “efecto de lomada” que se observa, y, por otro lado, mostrar como varía la estimación final del factor de inflación en función de estas variables, y por ende, evaluar si con otros parámetros mejora la subestimación del factor de inflación α para el caso del modelo imperfecto clásico.

Se toma como experimento de referencia el modelo Lorenz’96, y la estimación del factor de inflación α en el marco imperfecto con los parámetros clásicos, es

decir, utilizando como parámetro de forzado $F = 8$ para generar la verdad, y $F = 10$ para asimilar. Esto corresponde al Experimento 5.2.2. cuyos resultado de estimación de α con L96 se muestra en la Tabla 5.3.

En este experimento se variaron la cantidad de observaciones sintéticas (K), la cantidad de miembros del ensamble (N_e), y el valor donde se inicializa el algoritmo (α_0). Se tomaron las siguientes variaciones, y cada una fue realizada de manera independiente, dejando fijas las otras variables con los parámetros del experimento de referencia. Se varía la cantidad de observaciones en $K \in \{100, 500, 1000, 1500, 2000, 2500, 5000\}$, la cantidad de miembros del ensamble en $N_e \in \{50, 100, 500, 1000, 1500\}$ y el valor donde se inicializa el algoritmo en $\alpha_0 \in \{0.5, 0.8, 1.0, 3.0, 5.0\}$.

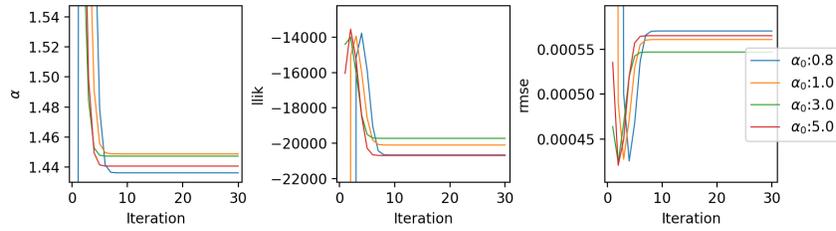
Se muestra en la Fig. 5.6 los resultados obtenidos. Se busca analizar si estas variables influyen para marcar este “efecto de lomada” que se observa en el experimento de referencia, y, por otro lado, mostrar como varía la estimación final del factor de inflación en función de estas variables.

Con respecto a la influencia de estas variables en el “efecto lomada” que se observa en la log-verosimilitud, y consecuente efecto en el RMSE, se puede observar que no se modifica con la variación del factor inicial, ni con la cantidad de observaciones, pero sí que disminuye a medida que aumenta la cantidad de miembros del ensamble.

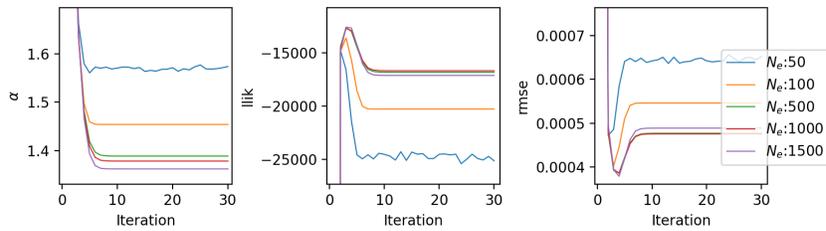
Se muestra en la Figura 5.6.a que al modificar el valor de inicialización del algoritmo, las estimaciones del factor de inflación α no dan variaciones apreciables (éstas son menores o del orden de 0.01), y no se observa algún patrón que indique que un valor inicial es mejor que otro, siempre en cuanto, el algoritmo converja.

Con respecto al número de miembros del ensamble, Fig. 5.6.b, se muestra que a medida que aumenta la cantidad de miembros del ensamble, la estimación del factor de inflación es menor. Es decir, al aumentar la cantidad de muestras, disminuye el valor del factor de inflación (en el orden de 0.2), ya que este no debe representar el error por muestreo, sin embargo, al estar asimilando con modelo imperfecto, necesita seguir siendo mayor a 1. Este comportamiento es el esperable, aunque se observa que aumenta el efecto lomada cuando disminuye el número de miembros del ensamble (por lo que la subestimación por el algoritmo es aun mayor en este caso).

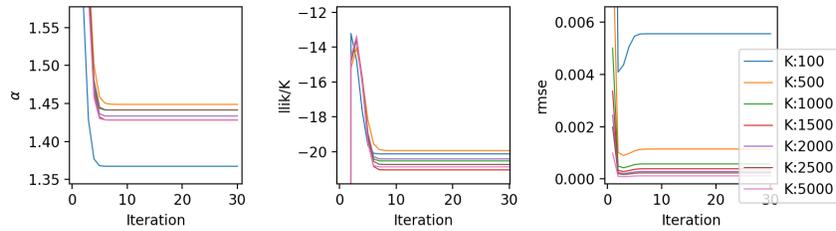
Por último, en la Fig. 5.6.b se muestra el experimento que varía la cantidad de observaciones sintéticas: para este caso en particular se grafica normalizada la log-verosimilitud. Se muestra que los valores de log-verosimilitud que se estiman son cercanos entre sí, para las distintas cantidades de observaciones y no hay alguna correlación entre aumentar la cantidad de observaciones y que aumente o disminuya la log-verosimilitud; sin embargo, si se observa en el RMSE y en la estimación de α que el caso de 100 observaciones es diferente a los demás, mostrando que hay un número de observaciones mínimas para el método, ya que con 100 observaciones no se cuenta con suficiente información para que el modelo pueda optimizar. Por otro lado, a partir de 500 observaciones hasta las 5000 testeadas, se muestra que las diferencias de estimaciones del factor son del orden del 0.01, similar a cuando se modifica la estimación inicial, y los valores



(a) Variando la inicialización del algoritmo



(b) Variando cantidad de miembros del ensamble



(c) Variando la cantidad de observaciones sintéticas

Figura 5.6: Experimento 5.2.4: Variando los parámetros del experimento de Estimación de α con EM utilizando el modelo L96 en un marco Imperfecto.

de log-verosimilitud y RMSE son similares. Esto muestra que estas diferencias están dentro de las variaciones del mismo error de estimación, pero no habría algún patrón al aumentar la cantidad de observaciones.

Como conclusión, en este capítulo se extendió el método EM para la estimación de los parámetros denominados β y α , se implementaron estas modificaciones y se testearon el rendimiento en diferentes marcos de experimentos imperfectos. Si se toma como referencia las estimaciones obtenidas con los métodos Grid Search y Nelder-Mead, se mostró que existe error en la estimación de estos parámetros con este método. En el caso del factor multiplicativo de la covarianza (β) se tiende a sobreestimar, y en el caso del factor de inflación (α) se tiende a subestimar. Se mostró que esta subestimación disminuye, comenzando a estimar mejor el algoritmo, cuando disminuye el error de modelo y error de muestreo a representar.

Se especula que el error de modelo sistemático, representado en los experimentos a través del error paramétrico, es el que produce un problema de convergencia en el algoritmo de EM secuencial. El error de modelo sistemático no es considerado por los análisis \mathbf{x}^a cuando solo se utiliza el filtro, por lo que existe una subestimación en cada ciclo del factor de inflación por no haber correcciones del factor debidas puramente a las observaciones. Por el contrario, cuando se utiliza el algoritmo EM tradicional con uso de observaciones futuras a través del suavizador permite corregir el error de modelo sistemático.

Capítulo 6

Conclusiones

En este trabajo se caracterizó a través de pruebas de concepto la subestimación de la covarianza del pronóstico en el filtro de Kalman por ensambles. Esta subestimación conlleva a la divergencia del filtro. Se mostró que en estos casos la divergencia ocurre al no lograr las matrices de covarianzas de error involucradas representar el error de modelo que se comete al realizar la asimilación con un modelo subrogado. Se mostró luego como la incorporación de parámetros de la covarianza adecuados en el sistema de asimilación pueden representar este error de modelo, y así lograr la no divergencia, y, como consecuencia, que la asimilación lograda sea un mejor estimador del estado real del sistema.

Se propuso utilizar y desarrollar técnicas de optimización que maximicen la verosimilitud para encontrar esos parámetros de la covarianza óptimos. Se trabajó, por un lado, con los algoritmos Grid Search y Nelder-Mead que optimizan una aproximación de la log-verosimilitud, y, por otro lado, se desarrolló e implementó extensiones del algoritmo EM, las cuales optimizan la log-verosimilitud completa por medio de una función intermedia.

El método Grid Search es en esencia un método de evaluación de la medida y permite obtener un estimador relativamente robusto. Además, al realizar evaluaciones de la medida, permite estudiar la geometría del problema y así evaluar si el valor óptimo es un punto óptimo global o no. Por otro lado, el algoritmo Nelder-Mead es un método numérico para minimizar una función objetivo en un espacio multidimensional que no requiere del gradiente, y optimiza basado solo en los valores de la función objetivo. Nelder-Mead es independiente de una malla prefijada a diferencia de Grid Search, pero si es dependiente de un valor inicial para realizar las iteraciones del algoritmo, y no es factible conocer si se está estimando un valor óptimo local o global, sin embargo, el método Nelder-Mead es más viable que el Grid Search para escalar la técnica a más dimensiones.

Por otro lado, el algoritmo EM es un método iterativo que utiliza una función intermedia, la cual aproxima de manera local la función objetivo a optimizar. De esta manera permite optimizar iterativamente la función log-verosimilitud explícita sin necesidad de realizar aproximaciones de ella. Se trabajó en la extensión de este método para los parámetros de la covarianza, ya que fue probado

en contextos de asimilación, y además, tiene la posibilidad de ser escalable a dimensiones altas de parámetros, por ejemplo, se podría estimar los parámetros de la covarianza a la vez que se estima valores de las matrices de error de modelo y/o de observación.

Como resultado de este trabajo se logró mostrar que es posible estimar parámetros óptimos para las matrices de covarianza a partir de un conjunto de observaciones distribuidas en el tiempo. Esto cobra relevancia al contrastar con la bibliografía, ya que se encuentran trabajos como el de Miyoshi (2011) [9], en el que se realizaron estimaciones del factor de inflación, pero de manera online, es decir, se realiza una estimación para cada tiempo.

En este trabajo se mostró que en casos de estudio de pocas dimensiones, la optimización de log-verosimilitud aproximada, por medio de los algoritmos Grid Search y Nelder-Mead, da buenos estimadores para los parámetros de la covarianza. No obstante también se deja evidencia de que, cuando hay presencia de error de modelo, el valor de la estimación de los parámetros de la covarianza, y la incerteza en la estimación de los mismos, depende de la muestra (miembros del ensamble) y del error de las observaciones.

Por otro lado, como resultado, también en casos de estudio de pocas dimensiones, se obtuvo que la extensión del algoritmo EM desarrollada e implementada para estimar los parámetros de la covarianza no logran buenas estimaciones de dichos parámetros en contextos de modelo imperfecto con error sistemático. Se tiende a sobreestimar el factor multiplicativo de la covarianza, y subestimar el factor de inflación. Se obtuvo también que a medida que disminuye ese error de modelo paramétrico, el algoritmo si logra converger a valores óptimos.

Con respecto al parámetro multiplicativo α , como hipótesis de que la técnica no estime los valores óptimos puede pensarse que, utilizando el EM secuencial no fue factible incluir al smoother en la deducción del estimador, el smoother permite corregir las estimaciones basados en observaciones futuras lo cual acota el efecto del error sistemático del modelo. Es decir que el estimador deducido por medio del EM secuencial no incluye en sí la información de las observaciones futuras, por ende, esto implica que si se esté alejando de la realidad el estimador no pueda captar esta divergencia. Una idea para deducir un estimador que incluya las observaciones futuras es utilizar un esquema del estado anterior. Como trabajo a futuro por ende puede trabajarse en que el estimador del parámetro multiplicativo incluya las observaciones.

Por último, también como trabajo a futuro, se deja el trabajar con modelos realistas de mayor dimensionalidad y con observaciones reales para validar o evaluar las técnicas desarrolladas, ya que esta tesina utilizó los modelos Lorenz 63 y Lorenz 96, que son modelos de pocas variables que se utilizan solo para el ámbito de estudio.

Capítulo 7

Anexo

7.1. Glosario

Se deja a continuación un breve resumen de los términos más usados en el sistema de asimilación con Filtro de Kalman por ensamble (EnKF), junto con una gráfica para facilitar su comprensión. Las ecuaciones y explicaciones detalladas se encuentran en la sección (2.2.6).

En la Fig. (7.1) se representa un ciclo de asimilación, el que corresponde entre el tiempo $t - 1$ y t . Se observa como el conjunto de análisis en el tiempo $t - 1$ evoluciona al tiempo t por medio de la función f , que representa un modelo matemático de predicción, o pronóstico, y q_n que representa una realización del error aditivo del modelo, es decir, de una realización de una $\mathcal{N}(0, \mathbf{Q}_t)$.

Itemizamos a continuación los términos más usados del sistema de asimilación en esta tesis, y la notación utilizada en cada caso.

- Análisis, ensamble de análisis: un ensamble de análisis es un conjunto de análisis. El análisis es el producto resultando de la asimilación de datos. Cada análisis es la combinación estadística de un pronóstico y una observación, donde cada fuente de información tiene su error.

Notación: $\mathbf{x}_{k,j}^a$ = es la partícula j del ensamble de análisis en el tiempo k .

- Covarianza del ensamble de análisis (o covarianza del análisis), error del análisis: es la covarianza que se calcula a partir de los miembros del ensamble de análisis. Se utiliza el término error de análisis para hablar de la variabilidad de este conjunto, para un tiempo dado.

Notación: \mathbf{P}_k^a es la covarianza del ensamble de análisis en el tiempo k (o covarianza del análisis).

- Observación, operador de observación, covarianza de error de observación, error de observación: Las observaciones en el tiempo t , \mathbf{y}_k , están relacionadas con las variables ocultas por medio del operador de observación \mathcal{H} , tal que $\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \epsilon_k$.

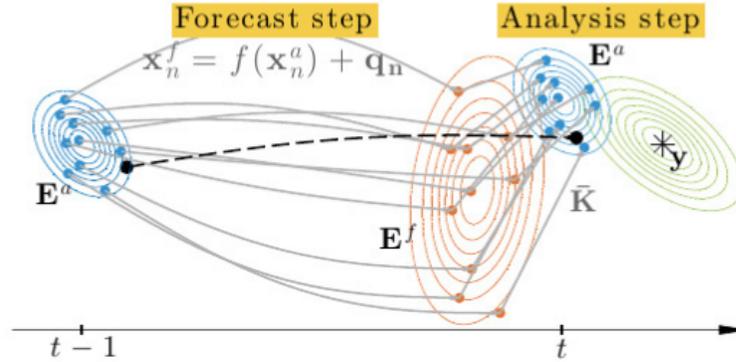


Figura 7.1: Representación de un ciclo de asimilación del filtro de Kalman por Ensamble. Se representa con azul los ensambles de análisis, donde cada punto representa un análisis, con naranja el ensamble de pronóstico, y con verde una observación.

ϵ_k es un error aditivo aleatorio de observaciones, que se asume gaussiano, con media cero, y matriz de error de covarianza \mathbf{R}_k , es decir, $\mathbf{Y}_k \sim \mathcal{N}(\mathcal{H}(\mathbf{x}_k), \mathbf{R}_k)$. Se utiliza el término error de observación para hablar de la matriz de covarianza del error de observación. En esta tesis se utilizó la matriz \mathbf{R} fija, independiente del tiempo k .

Notación: \mathbf{y}_k observación en el tiempo k , \mathcal{H} operador de observación, \mathbf{R} covarianza de error de observación.

- Pronóstico, ensamble de pronóstico, error de modelo, error aditivo del modelo: el ensamble de pronóstico en el tiempo t se genera a partir de evolucionar cada uno de los miembros del ensamble de análisis en el tiempo $t-1$ con un modelo matemático de predicción, $f_{k-1,k}$, y añadirle un error aleatorio tal que $\mathbf{x}_{k,j}^f = f_{k-1,k}(\mathbf{x}_{k-1,j}^a) + \eta_{k,j}$.

$\eta_{k,j}$ es un error aditivo aleatorio del modelo, que se asume gaussiano con media cero y matriz de error de covarianza \mathbf{Q}_k , es decir, $\mathbf{X}_k \sim \mathcal{N}(f_{k-1,k}(\mathbf{x}_{k-1}), \mathbf{Q}_k)$. En esta tesis se utilizó la matriz \mathbf{Q} fija, independiente del tiempo k . También se la llama a esta matriz error aditivo del modelo.

Notación: $\mathbf{x}_{k,j}^f$ es la partícula j del ensamble de pronóstico en el tiempo k , $f_{k-1,k}$ es el modelo matemático de predicción del tiempo $k-1$ al tiempo k , \mathbf{Q} matriz de error de modelo.

- Covarianza del ensamble de pronóstico (o covarianza del pronóstico): es la covarianza que se calcula a partir de los miembros del ensamble del pronóstico.

Notación: \mathbf{P}_k^f es la covarianza del ensamble de pronóstico en el tiempo k (o covarianza del pronóstico).

- Factor de inflación aditivo: es el parámetro multiplicativo de la covarianza de error de modelo. Es decir, multiplica a la matriz de covarianza \mathbf{Q} .

Notación: β es el factor de inflación aditivo, quedando la nueva matriz de covarianza de error de modelo $\beta\mathbf{Q}$.

- Factor de inflación multiplicativo: es el parámetro multiplicativo de la covarianza del pronóstico. Es decir, multiplica a la matriz de covarianza \mathbf{P}_k^f , en cada tiempo k . Se aplica actualizando los miembros del ensamble (eq. (2.25), sec. (2.2.8)).

Notación: α es el factor de inflación multiplicativo, quedando la nueva matriz de covarianza del ensamble del pronóstico $\alpha\mathbf{P}_k^f$.

- Error de pronóstico: el error de pronóstico involucrado en el ciclo de asimilación k proviene de dos fuentes: la matriz aditiva de error de modelo \mathbf{Q} , y la matriz de covarianza del ensamble del pronóstico \mathbf{P}_k .

Notación: $\alpha\mathbf{P}_k + \beta\mathbf{Q}$ es la matriz de covarianza de la predicción efectiva después de estas modificaciones con ambos factores de inflación.

- Error sistemático: en general en estadística se utiliza este término para referirse a aquel error que se produce de igual modo en todas las mediciones que se realizan de una magnitud. En esta tesis se utilizó para referirse al error generado por asimilar con un modelo diferente al utilizado para representar la realidad. Los experimentos denominados experimentos imperfectos tienen presencia de error sistemático.

7.2. Resumen de experimentos

Se deja a continuación un resumen de los experimentos realizados en cada capítulo, con sus objetivos, tipo de experimentos, modelo utilizado, y parámetros seteados tanto en la generación de las observaciones sintéticas, como en el proceso de asimilación, y parámetros utilizados.

Experimento	Objetivo	Tipo de experimento	Modelo	Parámetros en el proceso de generación de obs. sintéticas				Parámetros en el proceso de generación del filtro		
				Q	R	K	α	Q	Ne	α
3.1	Validar relevancia del factor de inflación	Modelo Perfecto	L63	0	1.5	1000	1	0	100	1
3.2			L96							1
3.3		Modelo Imperfecto	L63	0	1.5	1000	1	0	100	1
3.4										1.78
3.5										1
3.6										1.67

Figura 7.2: Experimentos capítulo 3

Experimento	Tipo de experimento	Objetivo	Modelo	Parámetro a estimar	Parámetros en el proceso de generación de obs. sintéticas				Parámetros en el proceso de generación del filtro		
					Q	R	K	α	Q	Ne	α
4.1.1	Modelo Perfecto	Validación de la técnica	L96	β	$\beta \cdot Q_f$ $\beta = 1.3$ $Q_f = 0.011$	0.5	1000	1	$\beta \cdot Q_f$ $\beta = [1 : 0.01 : 2]$ $Q_f = 0.011$	1000	1
4.1.2		Influencia del tamaño muestral				α					
4.2.1-a	Modelo Imperfecto	Estimación de parámetro	L63	α	0	0.5	1000	1	0	100	$\alpha = [1 : 0.01 : 2]$
4.2.1-b						1					
4.2.2			L96	α	0	0.5	1000	1	0	100	$\alpha = [1 : 0.01 : 2]$
4.2.3						1					
4.2.3	(α, β)	0	0.5	1000	1	$\beta \cdot Q_f$ $\beta = [1 : 0.01 : 2]$ $Q_f = 1$	1000	1	100	$\alpha = [1 : 0.01 : 2]$	

Figura 7.3: Experimentos capítulo 4

Experimento	Parámetro a estimar	Modelo	Tipo de experimento	Objetivo	Parámetros en el proceso de generación de obs. sintéticas				Parámetros en el proceso de generación del filtro		
					Q	R	K	α	Q	Ne	α
5.1.0	β	L96	Modelo Perfecto	Validación de la técnica	$\beta \cdot Q_f$ $\beta = 1.3$ $Q_f = 0.011$	0.5	1000	1	$\beta \cdot Q_f$ $\beta = [1 : 0.01 : 2]$ $Q_f = 0.011$	1000	1
5.1.2			Modelo Imperfecto	Estimación del parámetro		0					
5.2.1	α	L63	Modelo Imperfecto	Estimación del parámetro	0	0.5	1000	1	0	100	$\alpha = [1 : 0.01 : 2]$
5.2.2		L96	0	0.5	1000	1	0	100	$\alpha = [1 : 0.01 : 2]$		

Figura 7.4: Experimentos capítulo 5 - Parte I

Experimento	Parámetro a estimar	Modelo	Tipo de experimento	Objetivo	Parámetros en el proceso de generación de obs. sintéticas					Parámetros en el proceso de generación del filtro			
					Q	R	K	α	F	Q	Ne	α	F
5.2.3	α	L96	Modelo Imperfecto	Estimación en diferentes grados de error paramétrico	0	0.5	1000	1	8	0	100	$\alpha = [1 : 0.01 : 2]$	8 8.1 8.5 9 10

Figura 7.5: Experimentos capítulo 5 - Parte II

Capítulo 8

Bibliografía

- [1] E. Kalnay, *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.
- [2] F. Bouttier and P. Courtier, “Data assimilation concepts and methods march 1999,” *Meteorological training course lecture series. ECMWF*, p. 59, 2002.
- [3] A. C. Lorenc, “Modelling of error covariances by 4d-var data assimilation,” *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, vol. 129, no. 595, pp. 3167–3182, 2003.
- [4] R. E. Kalman, “A new approach to linear filtering and prediction problems,” 1960.
- [5] G. Evensen, “Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics,” *Journal of Geophysical Research: Oceans*, vol. 99, no. C5, pp. 10 143–10 162, 1994.
- [6] M. Katzfuss, J. R. Stroud, and C. K. Wikle, “Understanding the ensemble kalman filter,” *The American Statistician*, vol. 70, no. 4, pp. 350–357, 2016.
- [7] J. L. Anderson, “Localization and sampling error correction in ensemble kalman filter data assimilation,” *Monthly Weather Review*, vol. 140, no. 7, pp. 2359–2371, 2012.
- [8] T. M. Hamill, J. S. Whitaker, and C. Snyder, “Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter,” *Monthly Weather Review*, vol. 129, no. 11, pp. 2776–2790, 2001.
- [9] T. Miyoshi, “The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform kalman filter,” *Monthly Weather Review*, vol. 139, no. 5, pp. 1519–1535, 2011.

- [10] G. Desroziers, L. Berre, B. Chapnik, and P. Poli, “Diagnosis of observation, background and analysis-error statistics in observation space,” *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, vol. 131, no. 613, pp. 3385–3396, 2005.
- [11] T. M. Hamill and C. Snyder, “A hybrid ensemble kalman filter–3d variational analysis scheme,” *Monthly Weather Review*, vol. 128, no. 8, pp. 2905–2919, 2000.
- [12] A. M. Clayton, A. C. Lorenc, and D. M. Barker, “Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office,” *Quarterly Journal of the Royal Meteorological Society*, vol. 139, no. 675, pp. 1445–1461, 2013.
- [13] A. C. Lorenc, N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, “Comparison of hybrid-4denvar and hybrid-4dvar data assimilation methods for global nwp,” *Monthly Weather Review*, vol. 143, no. 1, pp. 212–229, 2015.
- [14] M. Bonavita, E. Hólm, L. Isaksen, and M. Fisher, “The evolution of the ecmwf hybrid data assimilation system,” *Quarterly Journal of the Royal Meteorological Society*, vol. 142, no. 694, pp. 287–303, 2016.
- [15] D. T. Kleist and K. Ide, “An osse-based evaluation of hybrid variational-ensemble data assimilation for the ncep gfs. part ii: 4denvar and hybrid variants,” *Monthly Weather Review*, vol. 143, no. 2, pp. 452–470, 2015.
- [16] M. Pulido, P. Tandeo, M. Bocquet, A. Carrassi, and M. Lucini, “Stochastic parameterization identification using ensemble kalman filtering combined with expectation-maximization and newton-raphson maximum likelihood methods,” *arXiv preprint arXiv:1709.07328*, 2017.
- [17] P. Tandeo, M. Pulido, and F. Lott, “Offline parameter estimation using enfk and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization,” *Quarterly Journal of the Royal Meteorological Society*, vol. 141, no. 687, pp. 383–395, 2015.
- [18] D. Dreano, P. Tandeo, M. Pulido, B. Ait-El-Fquih, T. Chonavel, and I. Hoteit, “Estimating model-error covariances in nonlinear state-space models using kalman smoothing and the expectation-maximization algorithm,” *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 705, pp. 1877–1885, 2017.
- [19] O. Cappé, E. Moulines, and T. Rydén, “Inference in hidden markov models,” in *Proceedings of EUSFLAT conference*, 2009, pp. 14–16.
- [20] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

- [21] R. Todling, “Estimation theory and foundations of atmospheric data assimilation,” *DAO Office Note*, vol. 1, p. 1999, 1999.
- [22] G. Burgers, P. Jan van Leeuwen, and G. Evensen, “Analysis scheme in the ensemble kalman filter,” *Monthly weather review*, vol. 126, no. 6, pp. 1719–1724, 1998.
- [23] G. Evensen, “The ensemble kalman filter: Theoretical formulation and practical implementation,” *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [24] A. Doucet, N. De Freitas, N. J. Gordon *et al.*, *Sequential Monte Carlo methods in practice*. Springer, 2001, vol. 1, no. 2.
- [25] P. Del Moral, A. Doucet, and A. Jasra, “Sequential monte carlo samplers,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.
- [26] P. Raanes, “Introduction to data assimilation and the ensemble kalman filter,” 2016.
- [27] J. S. Whitaker and T. M. Hamill, “Ensemble data assimilation without perturbed observations,” *Monthly Weather Review*, vol. 130, no. 7, pp. 1913–1924, 2002.
- [28] E. Cosme, J. Verron, P. Brasseur, J. Blum, and D. Auroux, “Smoothing problems in a bayesian framework and their linear gaussian solutions,” *Monthly Weather Review*, vol. 140, no. 2, pp. 683–695, 2012.
- [29] T. J. Sullivan, *Introduction to uncertainty quantification*. Springer, 2015, vol. 63.
- [30] A. Carrassi, M. Bocquet, A. Hannart, and M. Ghil, “Estimating model evidence using data assimilation,” *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 866–880, 2017.
- [31] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [32] R. H. Shumway and D. S. Stoffer, “An approach to time series smoothing and forecasting using the em algorithm,” *Journal of time series analysis*, vol. 3, no. 4, pp. 253–264, 1982.
- [33] E. N. Lorenz, “Deterministic nonperiodic flow,” *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
- [34] —, “Predictability: A problem partly solved,” in *Proc. Seminar on predictability*, vol. 1, no. 1, 1996.
- [35] M. Pulido and O. A. Rosso, “Model selection: using information measures from ordinal symbolic analysis to select model subgrid-scale parameterizations,” *Journal of Atmospheric Sciences*, vol. 74, no. 10, pp. 3253–3269, 2017.